ABSTRACT. A general characterization of functionalist theories of mind is offered and a number of issues are discussed which allow for alternative versions of functionalism. Some issues, such as the distinction between the implicit definition and partial specification views are of a general nature, while others raise questions more specific to functionalism, such as whether the relation between psychological and physiological properties is one of identity or instantiation. Section II attempts to undermine several arguments which have been offered to support the non-identity position. In the final section, the suggestions that the relevant notions of functional equivalence might be unpacked solely in terms of abstract automata features or entirely in terms of causal relations to nonintentionally characterized behavior are shown to be inadequate.

Among the theories of mind which have been proposed in recent years, a great many might be termed in one sense or another functionalist. The many proponents of functionalism have differed greatly both in their formulations of the thesis and about its implications. Functionalism has been put forward as an argument for the psychophysical identity thesis and as a refutation of that view. It has been taken by some to show the reducibility of psychology to physics, and by others as a basis for antireductionism. Still others have found in functionalism support for a disappearance theory of mental terms. Functionalism as so influential but varied a theory of mind deserves further clarification and examination. In section I of this paper, several distinctions among formulations of functionalism are drawn, while section II and III respectively focus on two particular issues critical to the overall functionalist approach.

Copyright, 1983, Philosophy Research Archives
FUNCTIONALISM

I.

Perhaps the most general feature of functionalist theories of mind is that they characterize particular types of psychological states (or properties) in terms of the roles played by those states within a structure mediating perceptual stimulation and behavior. A weak functionalist claim that at least in some cases behavioral and stimulus relations are relevant in explicating the nature of particular psychological states seems so obviously true as to be uninteresting. Surely any adequate account of desire, for example, will make reference to the roles played by desires in conjunction with beliefs in bringing about behavior directed at the desire's satisfaction. We recognize that how a person behaves at a given time is determined, at least in part, by his current and past psychological states and that his present psychological makeup, similarly, is influenced by his past psychological states and stimulations. Few would dispute that these interconnections are relevant to characterizing psychological state types.

However, controversy and interest rises when it is maintained that significant classes of mental states can be fully explicated solely in terms of such functional relations. The stronger functionalist claim is that type identity conditions for mental states can be specified purely in terms of their mutual interconnections and relations to stimulus conditions and behavior. The specification of a psychological state's role within a theory of behavior is held not merely relevant to its explication, but sufficient as well.

Arriving at a precise statement of the strong functionalist claim requires clarification at several points. In holding that psychological states are functional states, we need to specify which states we are discussing. Functionalism might be taken as a thesis about ideal psychological theories, a claim that the psychological states referred to by the correct, best, or any fully adequate theory of human psychology will be functionally individuated states. On such an account, psychological states might be functional in character, even though the states referred to by our present psychological theories were not. However, most functionalist claims have been made with respect to current psychological theories, especially those utilizing the vocabulary of common sense psychology. It has been argued that such states or properties as believing, desiring, and intending, as well as believing that P and intending that Q are purely functional in nature. Functionalism thus construed as a thesis about the character of familiar intentional states is independent of the ideal theory version of functionalism, though compatible with it. For one might maintain or deny that such states or properties as believing and desiring will be among those referred to by a correct or fully ade-
FUNCTIONALISM

A further clarification is needed, since even those versions of functionalism which agree that the states referred to by our present psychological theories are functional by nature, may differ concerning the extent to which our present theories succeed in characterizing those states. On the view of implicit definition, psychological state terms are completely defined by the laws and assertions of the theory in which they occur. What it is for a state to be a belief state, in the ordinary sense, is completely specified by the set of laws and assertions concerning beliefs contained in our common sense psychological theory. Conjoining the strong functionalist claim with the implicit definition position we arrive at the thesis that our present common sense theory of psychology provides us at least implicitly with sets of sufficient conditions for type individuating our common sense psychological states purely in terms of functional relations.

Alternatively, our present theory might be taken as only partially characterizing many of the types of psychological states to which it refers. The psychological state terms of the theory would not be defined by the laws of the present theory. Those laws would serve merely to secure the reference of the terms involved and to partially characterize the states referred to. Sufficient conditions for type individuating the psychological states involved would be available only after further development and elaboration of our theory. The functionalist might nonetheless maintain that when these conditions are discovered, they will turn out to be purely functional in nature. The partial specification view taken together with strong functionalism yields the thesis that the psychological states referred to by our present common sense psychological theory can in fact be type individuated purely on the basis of functional relations, though our present theory is not adequate in many cases to allow us to formulate those conditions.

II.

Having drawn our initial distinctions, we must turn to the fundamental question of what is involved in classifying a state or property as functional. Our first suggestion was that such states were characterized by their role in mediating stimulation and behavior. In this respect functionalism is a sophisticated offspring of behaviorism, which also stressed the primacy of such interconnection in explicating psychological state terms. The behaviorist project of providing explicit behavioral definitions for psychological states on a term-by-term basis was frustrated at least in part because very little, if any, human behavior is a result of some independent psychological state plus stimulus conditions. Rather, how a person behaves is a function of the
FUNCTIONALISM

complex state of a large number of his individual psychological states. Similarly, the determinants of whether a person will now be in or acquire a given psychological state will include not only his present stimulus conditions, but his past stimulations and psychological states as well. Taking the holistic nature of psychological explanation seriously, the functionalist attempts to explicate the nature of particular psychological states within the context of an overall theory of human behavior and internal functioning. For a state to be a psychological state of a particular type is for it to occupy a particular location within a specified network of mutually interrelated states, behaviors, and stimuli.11

While the emphasis on the holism of the mental distinguishes functionalism from behaviorist theories, it is the claim that psychological properties are in some interesting sense non-structural which separates functionalism from classical statements of the psycho-physical identity thesis. Identity theorists had raised the possibility that being in such-and-such a particular mental state might turn out to be straightforwardly physiological.12 Some functionalists have found such conjectured type identities incompatible with their general functionalist approach. For on such a view, a person has a particular psychological property F just if he has a set of properties which are mutually interrelated to themselves, behavior, and stimuli in some specified way R.

The psychological property F is not to be identified with the set of properties which are related in way R, but rather with the property of having properties which are related in way R. Psychological properties, involving as they do quantification over properties are second order properties. Thus, it is argued that they cannot be identified with any straightforwardly physiological properties which are correctly understood as first order structural properties.13

The relation between physiological and psychological properties is held to be one of realization or instantiation, not identification. This construal of functional properties as second order properties raises several issues concerning functionalism which deserve critical examination.

If a given psychological property is to be understood as the property of having a set of first order structural properties which are related in way R, we need to explicate more fully the structural/functional distinction and to explain what sorts of relations and descriptions can be utilized in specifying the corresponding R relations. The first task occupies the remainder of this section, and the latter is the subject of section III.

The alleged functionalist refutation of the identity thesis rests on a characterization of psychological proper-
ties as functional and of physiological properties as structural. A difficulty immediately arises since states or properties cannot be characterized as functional or structural simpliciter; such distinctions can only be made relative to a level of abstraction. For what counts as a functional property of a system described at one level of abstraction may count as structural when described at another. Recall that a functional property is supposedly a property which a system has purely in virtue of its input/output relations.

The system's structural features are relevant to determining its functional properties only insofar as a given feature affects its input/output relations. Thus two systems having widely divergent structures nonetheless may have the same functional properties; i.e., they may be functionally equivalent systems in virtue of exhibiting the same input/output relations. But at what level of abstraction are we to describe the inputs and outputs involved? Our choice of a vocabulary for describing the relevant inputs and outputs will be a major factor in determining when two systems exhibit the same input/output relations.

Imagine two devices $S^1$ and $S^2$. $S^1$ takes as inputs random strings of words printed on a tape and produces as outputs tapes printed with the alphabetized string of the words printed on its immediately prior input. $S^2$ takes as inputs random strings of words spelled out in metal printer's type and produces as outputs strings of metal type spelling out the alphabetized string of words from its input. If we describe the inputs and outputs of $S^1$ and $S^2$ as strings of words, then $S^1$ and $S^2$ exhibit the same input/output relations, and they are functionally equivalent. However, if we describe their inputs and outputs as tapes and strings of metal or type, they are clearly not functionally equivalent; indeed, they exhibit no common input/output relation.

Thus, one might suppose that being an alphabetizer is a functional property, while being a tape printer and being a type rearranger are structural properties. Any such suggestion would be mistaken, for consider a third device, $S^3$, which like $S^1$ takes tapes of printed word strings as inputs and produces tapes of those words alphabetized as outputs. $S^3$ produces its outputs by erasing and reprinting on its input tape, while $S^1$ prints its alphabetized string in upper case symbols as a new double sized tape, and destroys its input tape. If we include reference to the particular tape tokens and to the sizes of the tapes and symbols in our input/output descriptions, $S^1$ and $S^3$ will clearly not be functionally equivalent. Similar results would follow if we decided to count $S^3$'s erasing activity and $S^1$'s printing of a new tape as among the output activities of those systems.

Any judgment that a given property is functional in
FUNCTIONALISM

nature or that two systems are functionally equivalent must always be relativized to a level of abstraction. The claim must be made with respect to a specification of what is to be counted as an input or an output, and of the vocabulary to be used in describing the inputs and outputs involved.

Thus the structural/functional distinction does not provide any absolute division of properties into distinct ontological categories. What is functional when described in one way may be structural when described in another. The distinction must always be drawn relative to a somewhat arbitrary scheme dividing function from underlying structure. The decision one must make is upon a level of abstraction, in the sense that it is a decision about the sorts of differences between systems from which one wishes to abstract in describing inputs or outputs and in delimiting the notion of functional equivalence. One decides what sorts of differences between A and B may be ignored in determining whether or not A and B function in the same, i.e., equivalent, ways. The question of which differences are differences of function and which mere differences of structure is always relative to the investigation at hand and includes a not undesirable but unavoidable element of choice. Thus, the claim that psychological properties are functional in nature must involve an implicit specification of a level of abstraction relative to which the claim is made.

The need to specify a level of abstraction in drawing the structural/functional distinction is acutely evident with respect to the properties of concern to the functionalist. For while physiological properties or states might be thought of as underlying structures realizing functionally characterized psychological states, they may also be described in physiological terms which themselves allow of functional interpretation. In describing a neuron as having a differential excitation threshold, variable recovery period or specific background firing rate, we ascribe to it properties which seem clearly functional and of higher order.

If physiological properties and other physical properties of many sorts can be all interpreted as functional properties, then the functional nature of psychological states cannot be taken as evidence against the identity thesis. Moreover, the very thesis of functionalism itself is in danger of losing interest, for its appeal lay in picking out some supposedly distinctive characteristic of psychological properties, while it now appears that many non-psychological properties show those characteristics well.

One might attempt to reformulate the nonidentity claim in a more comparative way while conceding that one cannot in any absolute sense classify physiological and psychological properties as respectively first and second order properties; one might nonetheless maintain that psychological
FUNCTIONALISM

properties are properties of an order higher than the order of physiological properties. Properties of both sorts are of higher order, in the sense of not being first order properties, but psychological properties might be taken to be of a comparatively higher order than physiological properties. This judgment, though comparative, would be made with respect to some absolute ordering of properties. This suggestion does not appear very promising as it presupposes some general scheme according to which we might hierarchically order properties of very diverse sorts, drawn from disparate theories. A thoroughgoing reductionist and firm believer in the overall unity of science might believe that such a hierarchy was constructible building from a base of simple primitive first order physical properties. However, such hopes seem unlikely to be realized. In addition to the difficulties which attend so radical a reductionist view, there is the problem that the properties referred to by the theories which ascend by many branches from this primitive base are not likely to nest in any neat way allowing themselves to be uniquely ordered. Branches are likely to diverge and merge in diverse and subtle ways, providing no single and non-arbitrary way of measuring a theory's or property's complexity. Thus, as long as claims about the orders of psychological and physiological properties are understood in terms of some absolute ordering, no such claim, comparative or otherwise, can provide much support for the nonidentity thesis.

A more relativistic approach to the argument might be tried. Recognizing that any judgment of even the comparative order of properties must be made relative to a specified level of abstraction, it might be argued that there is a scheme of abstraction and description which is particularly appropriate for psychology, and that with respect to the levels of abstraction built into that scheme psychological properties are of a higher order than physiological properties. It might even be maintained that relative to this scheme of abstraction, physiological properties come out as first order properties, while psychological properties are properties of higher order. It would then be possible to argue that relative to this scheme, physiological properties are structural, while psychological properties are functional and nonstructural, supporting the nonidentity thesis.

However, unless strong independent grounds can be given for accepting such a scheme, such a proposal simply begs the question at issue. The suggestion is that we draw the structural/functional distinction relative to a level of abstraction such that physiological properties fall on the structural side of the distinction, and psychological properties on the functional side. But such a choice of level presupposes that no psychological properties are physiological in nature; it involves a prejudgment that none of the properties referred to by our psychological terms are physiological properties. Thus, unless the choice of an approp-
FUNCTIONALISM

riate level can be independently justified, the argument above can provide at best circular support for the nonidentity thesis.

Some functionalists have attempted to justify their claim that psychological descriptions ought to abstract from matters of physiology by arguing that creatures radically different from humans in physiology might nonetheless share many of the psychological properties of humans. Suppose we discovered intelligent creatures on the third planet of Alpha Centauri. In the presence of noxious stimulation, they exhibit escape behavior and a great deal of agitation, which increases with the intensity of the noxious stimulation. The Centaurians take steps to avoid such noxious stimuli, and they comment on their condition during the periods of agitation as horrible and awful. The Centaurians, however, are very unhuman-like physiologically. Their body chemistry is silicon based and they lack any analogues of neurons. The physical determinants of their behavior are very unlike those of our nervous systems. The functionalist claim is that these facts about Centaurian physiology would not justify us in denying that Centaurians have or feel pain. Thus, the property of having a pain cannot be identified with any property of human-like physiology (e.g., the property of having one's c-fibers firing).

However, if the story of the Centaurians is varied slightly, the issue may no longer seem so clear. Imagine that we do not arrive at Centauri until 2577, and during the intervening 500 years we develop a very sophisticated and detailed theory of pain. This future pain theory might be one which incorporates the physiological nature of pain in a quite central way. The physiological aspects of the theory would not merely explain how the phenomenon of pain is realized by human physiological systems; rather, they would provide a fuller understanding of the phenomenon of pain itself. New behavioral, emotive, and perceptual features of pain and its relation to stimuli would be uncovered by physiological investigation. Conjecturing such a theory is not fanciful, as theories of this general sort are put forward by pain researchers today. In explaining the complex physiological phenomenon of pain, these researchers have appealed to the existence of such things as self-exciting neural loops, spinal gating mechanisms, a central biasing mechanism, and the ratio between small and large nerve fiber firing rates. These physiological features are not involved merely to explain the realization of a relatively complete functional account of pain. The spinal gating and central biasing mechanism, for example, involve complex and previously unrecognized relations between the intensity of noxious stimulation and the perceptual and emotive dimensions of pain. Another physiological hypothesis attempts to account for the disruptive character of pain by showing that pain perception is associated with (indeed perhaps identical with) disruption of patterned thalamic-cortical activity.
FUNCTIONALISM

The physiological aspects of these current theories, like those in our supposed 2577 theory, are intended as at least partial accounts of what the phenomenon of pain is, not merely as accounts of how it is realized. What then would be our judgment concerning the Centaurians in 2577?

It seems that we might very well conclude that they did not have pains, though there is some property G which they do have, and being G is in some respects like having a pain. In light of our 2577 theory of pain, the Centaurians' nonhuman physiology precludes their having pains. This would not constitute a denial that they have certain awful experiences which they attempt to avoid. The claim is only that in the envisioned case, the identity conditions for being the same psychological property as having a pain have turned out to include a number of physiological factors as we developed our theory of the complex phenomenon of pain.

The example is not intended to show that the conditions for having a pain will in fact turn out to include physiological factors, but only that in plausible ways they might. Consider an analogous case in genetics. The term "gene" was originally introduced with respect to the functional role of genes in explaining actually observed phenomena, and classical genetics continued to expand the theory of genes along generally functional lines. In recent decades, biochemical researchers with respect to DNA have led to the identification of genes with portions of DNA molecules. But what if we discovered Centaurians who conformed to the laws of classical genetics, but who lacked any DNA molecules? How would such a situation be correctly described? Would we say that Centaurians had genes but their genes were not identical with DNA? Or would we say that Centaurians lacked genes, but had some structures which played a role very similar to that played in humans by genes? We must choose the first description if we are to treat genes as the functionalist would have us treat pains, yet there seem strong considerations against doing so. Most important is that the term "genes" acquired its reference with respect to a particular set of phenomena of humans and other earth creatures. While I do not believe that either of the possible descriptions of Centaurian heredity is clearly correct, this demonstrative (earth-fixing) element in the reference of "genes" tends to support the view that in using "genes" we referred to DNA molecules.

The parallel view that Centaurians do not have pains, but have only some related property is especially plausible if one maintains that there is also a demonstrative element present in the reference fixing of such psychological terms as "pain." One may hold that "pain" refers to some complex property (or phenomenon) of human beings, and that investigations of human neurophysiology are relevant to explicating its nature. The merits of this position should be clearer if we distinguish between two sorts of psychological enter-
FUNCTIONALISM

prises. One takes as its task the giving of some general theory with concepts and laws that apply to all sentient creatures. The other has a more anthropocentric goal, that of giving an account of human behavior and mental functioning, perhaps extending its scope to giving related accounts for creatures relatively akin to humans, such as monkeys, dogs, or rats. A theory of this latter sort would be a theory of the particular psychological processes which are the determinants of human behavior, and thus the referents of many of its psychological terms would be fixed with respect to the human case. Moreover, it seems likely that at least some physiological features, such as the spinal gating mechanism, might have to be included among the psychologically significant determinants of human behavior discussed by such a theory. Since our common sense psychological theories, as well as most of the theories offered by researchers, appear to be theories of this second sort rather than general theories of sentience, it seems reasonable to hold that physiological factors may be relevant to explicating the nature of some of the states and properties referred to, by such familiar psychological terms as "pain" or even "belief."

Most importantly, there appears to be no overriding reason for adopting the nonidentity position and restricting physiological factors to the level of realization and instantiation. We would regard as confused, or at least of doubtful value, any attempt by contemporary geneticists to distinguish between genetic properties and properties which merely realize genetic properties. Genetics aims at giving an overall account of the structures and processes involved in the development, utilization, regulation, and transference of hereditary information. Investigating the chemistry of DNA is a key part of discovering the nature and organization of human heredity. If the aim of our psychological theory is analogously to give an overall account of the psychologically significant determinants of human behavior, it would make no more sense for the psychologist to restrict physiological factors to the level of instantiation than it would the analogous restriction in genetics.

III.

Having concluded that there is no a priori reason to draw the structural/functional distinction so as to support the nonidentity thesis, we are left with the task of attempting to clarify the nature of the functional equivalences which are supposed by the functionalist to type-individuate psychological states. If functionalism is taken as the claim that to have a psychological property \( \Phi \) is to have properties \( P^1 \), \( \ldots \), \( P^n \) which are interrelated in way \( R \), we must say something about the nature of the \( R \) relations, which are associated with the various psychological state types. For it is these second order relations which deter-
mine the nature of the functional equivalence relation and it is the functional equivalence relation which in turn type-individuates $\Psi$-states.

Thus the sorts of descriptions and distinctions built into our specifications of the various $R$-relations will be the ones relevant to explicating the notion of functional equivalence. For example, if the $R$-relations included no reference to physiological features, then the corresponding notion of functional equivalence would similarly abstract from matters of physiology.

No attempt will be made here to specify any particular $R$-relation in detail, a task which might require us to provide a substantial psychological theory. Rather it will be argued that there are certain conditions which must be satisfied by an acceptable account of these $R$-relations, and that these conditions cannot easily be satisfied by proposals which attempt to specify the $R$-relations purely in terms of the features of abstract automata or purely in terms of causal relations to non-intentionally described behavior.

To specify conditions of adequacy for a theory of $R$-relations, we must understand the intended role of such a theory. There are three sorts of items which must be distinguished: our common sense psychological theory (or theories), functional theories which are intended to provide type-identify conditions for the properties and states referred to by our common sense theory, and the persons who are supposedly described by these two sorts of theories. The functional theories stand as it were between our common sense theory and the persons it supposedly describes. The three items might be pictured as levels.

Levels:

1. COMMON SENSE
2. PSYCHOLOGICAL THEORY
3. FUNCTIONAL THEORY
4. PERSONS

If the functional theory is supposed to provide type-identity conditions for the psychological states referred to by our common sense theory, it must satisfy constraints imposed both from above and below. If various states of the functional theory are to be identified with the states referred to by our common sense theory, then the relations between those functional states will have to be analogous in important ways to the interstate relations specified (or implicit) in our common sense theory. This condition of significant analogy is required to secure that reference is preserved, that the state terms of the functional theory
FUNCTIONALISM

refer to the same states as the state terms of our common sense theory. The functional theory need not and is unlike­ly to be a perfect analogue of our common sense theory, but it must preserve many significant features of the common sense theory if it is to license claims of identification, rather than elimination and replacement. Constraints are also imposed from below. Our common sense psychological theory is intended to describe persons and their behavior. If various states in our functional theory are to be identified as the referents of common sense state terms, it must be that the person described by the common sense theory actually realize those functional states. Recall that the version of functionalism we are discussing takes the claim of functionalism as a claim about the nature of the states referred to by the terms of our common sense theory. Thus it requires that we be realists about the states of common sense psychology. Persons must actually have states of the kinds referred to by common sense psychology, and no functional states can be identified as states of the kinds referred to by our common sense theory unless the persons described by the common sense theory do indeed have those functional states. The task of the functionalist is thus to discover and specify a set of functional states which persons actually have and which are interrelated in ways which are significantly analogous to the interrelations between psychological states specified (or implicit) in our common sense theory.

This latter requirement of significant analogy between the functional theory and common sense psychology constrains the class of acceptable functional theories and limits the sorts of R-relation which will be adequate for specifying any such theory. For example, suppose a certain degree of rational interconnectedness is a significant feature of the states described by common sense psychology. Then the states of any acceptable functional theory must be interconnected in roughly analogous ways. (There must be an approximate isomorphism between the two theories.) The interstate connections specified by the functional theory must mirror the rational interconnections of the common sense theory. Accordingly, the (set of) R-relation used to specify the relations between states of the functional theory must be adequate to specifying an analogue of rational interconnectedness. Thus if the claim is made that an adequate functional theory can be specified in some particular vocabulary, say in terms of the features of abstract automata, it must be shown that we can specify an analogue of rational interconnectedness purely in terms of such automata features. To the extent that we cannot do so, any automata model would fail to mirror the interstate relations implied in our present psychological theory.

Let us look at these issues in slightly greater detail. Any adequate automata theory of belief would likely include some machine analogue of reasoning. For the ways in which a
person's beliefs set is modified across time by reasoning is an important element in our ordinary theories of belief. Thus, it may be useful to consider some of the difficulties which arise in trying to construct an automaton model of reasoning. Consider a simple information-processing model consisting of a processor and memory.21

The memory will contain at a given time a finite number of stored symbol structures which are the automata analogous of (believed) propositions. The processor given any symbol structure as input, determines the subsequent state of the memory as a function from the input and immediately prior state (that is, total content) of the memory. Thus the processor is really a mathematical entity: a function from a self-ordered pair consisting of symbol structures (inputs) and sets of symbol structures (antecedent memory contents) onto a set of sets of symbol structures (the subsequent memory contents).22 The modifications of the systems memory are intended to correspond to reasoned changes in a person's belief set over time.

Interpreting the automaton as a reasoner would involve setting up some correspondence between the symbol structures of the automaton and beliefs (or propositions). A condition of adequacy on such an automaton model would be that under the intended correspondence, the set of transitions determined by the processor approximates the set of belief transitions our psychological theory would attribute to the typical reasoner (or approximates the belief transitions found by experimental investigation).23

Some rational connections would be easy to build into the processor, especially those concerning simple formal relations. We could include, for example, a machine analogue of modus ponens. For a given proposition A, let C(A) be the symbol structure analogue of A under the specified correspondence. We would then build our processor so that if the system contained either member of the pair C(A),C(A→B) in its memory and was given the other member as input, it would add C(B) to the memory as well. By using such formal rules we might build into our processor analogues of many of the patterns of deductive inference employed by the typical
FUNCTIONALISM

reasoner. But there seem to be obvious limits to this strategy. A large portion of the evidential and inferential connections which we expect the average reasoner to recognize and act upon are not deductive relations; a great many are not so straightforwardly formal. If Jones believes his wall mirror is rectangular, he almost surely believes that the wall mirror is not round. And if he believes that the mirror remains in position on the wall, he most probably believes that it is securely fastened to the wall by hooks, screws or nails. It will be much more difficult, if not impossible, to capture these nonformal relations by rules specified purely in terms of operations on uninterpreted symbol structures. This would especially be the case with respect to the multitude of obvious inductive consequences which we expect the average reasoner to draw. Nor is the fact that believers make such inductive inferences an inessential feature about believing. For surely a person cannot believe a proposition unless she understands to a fair extent what the proposition means. But understanding a proposition cannot be separated from recognizing its interconnections with other propositions. To suppose that a person might believe a proposition without even implicitly recognizing any of its connections to other propositions is not merely implausible, it is incoherent. (Just as a sentence only has meaning within the context of a language, so a particular psychological state can only be a belief within the context of an interconnected system of beliefs.)

Recall that the condition of minimal adequacy proposed for reasoning automata requires that the transitions between sets of stored symbol structures mimic the average reasoner's belief transitions. But it is very difficult to see how we could build into the automaton's processor a set of rules for manipulating formally characterized symbol structures which would be sufficient to model the transitions that reasoners make on the basis of the nonformal and inductive relations between propositions.

Nothing said here precludes the possibility of giving an adequate functional account of intentional states purely in terms of the features of abstract automatons. Nor is that possibility one which I would want to rule out at this earlier stage of theory construction. Still the fact that such a project would apparently require giving some automaton account of rationality and formalizing a good bit of inductive logic, makes it seem considerably less promising (than might have been supposed).

Thus far, our discussion of rational correctedness has concerned only the relations which must obtain within one's belief set. Meaning and representation, however, concern not only relations among symbols (symbol structures) but also relations between (those) symbols and the world. And with respect to intentional systems, these symbol/world relations will again involve considerations of rationality.
Believing a proposition P involves, at least in many cases, having various complex behavioral dispositions. The cases which are of the greatest importance in characterizing the relations between intentional states and behavior are those in which a person’s intentional states serve as the reasons for her behaving as she did. These are typically cases in which the agent acts in a way which is likely to realize her goals on the assumption that her beliefs are true. Of course, much more is involved in acting on a reason; the nature of the agent's reasoning, weighing of desires, attention, judgments of probability, and causation are all relevant. Yet for present purposes only one feature need concern us. In explaining an agent's behavior by appeal to her reasons for action, we are constrained to attribute to her a set of beliefs, desires, preferences and intentions which is generally consistent and coherent. Moreover, it must be possible to characterize the behavior to be explained under some description plausibly intended by the agent, such that the imperative to so behave can be seen as the conclusion of a valid practical inference having as premises beliefs and desires attributed to the agent. Thus an automaton analogue of our psychological theory must mimic not only the rational interconnectedness among intentional states, but also the rational connections of those states to behavior. The automaton program is confronted with the additional task of having to specify some analogue of sound practical reasoning purely in terms of rules defined with respect to uninterpreted symbols, making the promise of that program very small indeed.

These behavioral relations raise another issue worthy of at least brief attention here. Some functionalists have argued that psychological states are type-individuated by their position within a network of causally interconnected states, stimuli, and behavior. Such versions of functionalism explicate the individuating R-relations primarily in terms of causal connection, rather than by reference to abstract machine features as in automaton versions of functionalism. Of particular interest with respect to causal versions of functionalism is the matter of behavior characterization, that is, the vocabulary employed in describing the behavioral elements in the functional networks. For if the proponent of some version of causal functionalism also attempts to specify the relevant behaviors in nonintentional terms, the claim amounts to a form of reductionism. The alleged reduction would be of the intentional to the nonintentional, explicating the nature of intentional states in terms of causal networks containing only nonintentional elements. The only elements in the relevant causal networks other than nonintentionally specified behaviors and stimuli would be internal place holders defined purely in terms of their causal relations.
Thus the claim that being a psychological state of a given type is to occupy one of the internal positions within such a network is reductionist in nature.

This sort of functionalism seems implausible as it is unlikely that we would be able to model the sorts of relations between intentional states and behavior implicit in our current psychological theories if restricted to only nonintentional descriptions of behavior. To show this, it is important to be clear about what is meant by a nonintentional description of behavior. Assuming we have some adequate idea of what counts as an intentional state, a nonintentional description of behavior is simply one which does not carry any entailments that the person so behaving is (or has been) in any particular intentional state (or states). "Stand up" would thus be a nonintentional description, while "promises to come" would surely be intentional.

As noted above, the behavioral relations which are most important in explicating the nature of intentional states are those in which the intentional states serve as an agent's reasons for behaving as she did. In such cases, attributing certain beliefs and desires to the agent will rationalize or explain her behavior only if the behavior is suitably described. The agent's behavior must satisfy or have been likely to satisfy some description which the agent intended her behavior to satisfy. Thus in attempting to explain the agent's behavior by recourse to her reasons for action, we must describe her behavior in a way which establishes the link between that behavior and the intended description which we attribute to her on the basis of her beliefs and desires.

In many cases, the description which the agent intends her behavior to satisfy will be intentional. Thus it would be very difficult, if not impossible, to explain behavior by reference to the agent's reasons for action while utilizing only nonintentional descriptions of behavior. Consider some examples. One might on occasion simply have a desire to move one's hand. If so, one's behavior could be described simply as moving one's hand and rationally explained by reference to the relevant desire. However, in most common cases of hand motion, one's intention is more likely to be something like shifting into third gear or leaving a message about defrosting dinner. To explain the agent's action and explicate the relation between her intentional states and behavior, we must describe the behavior with respect to her beliefs and desires.
FUNCTIONALISM

intentions, and in cases like the leaving of a message, it seems we must employ clearly intentional descriptions. In describing the agent's behavior as an act of communication, we attribute to her a wealth of intentional states. As such cases are far too pervasive and play too important a role in our psychological theory to be ignored, it appears unlikely that any adequate version of causal functionalism could be given which utilized only nonintentional descriptions of behavior. Thus whatever the difficulties and merits of causal functionalism in general may be, the issue of behavior specification greatly diminishes the prospects for any of its reductionist versions. Interestingly, the feature which undercuts the reductionist thesis is the same one which earlier defeated the reductionist program of radical behaviorism: the holistic nature of the intentional. The behaviorist project was frustrated by the interconnectedness of intentional states, which made it impossible to give analyses of those states which broke outside the circle of the intentional. What is brought out by the discussion above is that the circle of the intentional includes not only the interrelated internal states, but many of the relevant behaviors as well. For in many cases the analyses of intentional states describe types of behavior which cannot themselves be analyzed without reference to other intentional states. It appears that neither intentional attitudes nor intentional behavior can be explicated independently of one another.

Though the basic functionalist program remains intact,28 the specific suggestions that the functional equivalence relation might be unpacked solely in terms of abstract automata features or in terms of causal connections to nonintentionally characterized behavior must be rejected as inadequate.

FOOTNOTES


6. I think it is not unreasonable to consider Richard Rorty's views about intentional states as functionalist, though anti-mentalistic. See Rorty, R., Philosophy and the Mirror of Nature (Princeton, NJ, Princeton University Press, 1979), especially Chapter 2; Dennet also seems at times to imply that an adequate functionalist psychology will leave no room for mental states as we normally think of them. See the first chapter of Content and Consciousness and the essay, "Why You Can't Make a Computer that Feels Pain," in Brainstorms, 190-229. In as far as Quine is a functionalist rather than a behaviorist, he might also be classed in this group.


8. See especially Lewis, D., "Psychophysical and Theoretical Identification."


10. Some similar but not equivalent distinctions are drawn by Ned Block between what he calls "functionalism" and "psycho-functionalism" in his article, "Troubles with Functionalism," in Savage, W., (ed.), Perception and Cognition Issues in the Foundations of Psychology—Minnesota Studies in the Philosophy of Science, Vol. IX (Minneapolis, MN, University of Minnesota Press, 1979), 261-326. The present distinction concerns only different views about how the reference of psychological terms gets fixed while Block's distinction concerns also the kinds of relations which get built into the functional specification (e.g., physiological information in the psycho-physical cases).

11. The holistic motivation for functionalism is perhaps clearest in Grice, H.P., "Method in Philosophical Psychology."

13. As by Putnam in "The Nature of Mental States;" by Fodor in Fodor, J.A., Psychological Explanation (New York, Random House, 1968); and by Block and Fodor in "What Psychological States are Not."

14. These sorts of worries were first raised in Kalke, W., "What's Wrong with Fodor and Putnam's Functionalism," Nous, Vol. III (1969), 83-93.

15. Such a view is offered by Putnam in "On Properties."


18. The point of these examples is to turn Putnam's twin earth examples and general theory of reference back against his own claims about the necessarily functional character of pain. Given our ignorance about the nature of pain, it seems impossible to legislate in advance what sorts of considerations determine kinds in these areas. What we choose to treat as a kind may also depend upon our pragmatic and explanatory interests.

19. For a contrary view about the relation between classical and biochemical genetics see the first chapter of Hull, D., Philosophy of Biological Science (Englewood Cliffs, NJ, Prentice-Hall, 1974).


26. David Lewis, David Armstrong, and Gilbert Harman could all be con-
sidered advocates of such a view.

27. Here David Lewis must be excluded since he allows intentional entities such as propositions and attributes to serve as elements of the functional networks. For a difficulty raised by such an approach see Field, H., "Mental Representation," Erkenntnis, Vol. 13 (1978), 9-61.