4 Troubles with Functionalism

1.0 Functionalism, Behaviorism, and Physicalism

The functionalist view of the nature of the mind is now widely accepted.[1] Like behaviorism and physicalism, functionalism seeks to answer the question ''What are mental states?'' I shall be concerned with identity thesis formulations of functionalism. They say, for example, that pain is a functional state, just as identity thesis formulations of physicalism say that pain is a physical state.

I shall begin by describing functionalism, and sketching the functionalist critique of behaviorism and physicalism. Then I shall argue that the troubles ascribed by functionalism to behaviorism and physicalism infect functionalism as well.

One characterization of functionalism that is probably vague enough to be acceptable to most functionalists is: each type of mental state is a state consisting of a disposition to act in certain ways *and to have certain mental states*, given certain sensory inputs and certain mental states. So put, functionalism can be seen as a new incarnation, of behaviorism. Behaviorism identifies mental states with dispositions to act in certain ways in certain input situations. But as critics have pointed out (Chisholm, 1957; Geach, 1957; Putnam, 1963), desire for goal G cannot be identified with, say, the disposition to do A in input circumstances in which A leads to G, since, after all, the agent might not *know* that A leads to G and thus might not be disposed to do A. Functionalism replaces behaviorism's ''sensory inputs'' with ''sensory inputs and mental states''; and functionalism replaces behaviorism's ''dispositions to act'' with ''dispositions to act and have certain mental states.'' Functionalists want to individuate mental states causally, and since mental states have mental causes and effects as well as sensory causes and behavioral effects, functionalists individuate mental states partly in terms of causal relations to other mental states. One consequence of this difference between functionalism and behaviorism is that there are possible organisms that according to behaviorism, have mental states but, according to functionalism, do not have mental states.

So, necessary conditions for mentality that are postulated by functionalism are in one respect stronger than those postulated by behaviorism. According to behaviorism, it is necessary and sufficient for desiring that G that a system be characterized by a certain set (perhaps infinite) of input-output relations; that is, according to behaviorism, a system desires that G just in case a certain set of conditionals of the form ''It will emit O given I'' are true of it. According to functionalism, however, a system might have these input-output relations, yet not desire that G; for according to functionalism, whether a system desires that G depends on whether it has internal states which have certain causal relations to other internal states (and to inputs and outputs). Since behaviorism makes no such ''internal state'' requirement, there are possible systems of which behaviorism affirms and functionalism denies that they have mental states.[2] One way of stating this is that, according to functionalism, behaviorism is guilty of *liberalism*—ascribing mental properties to things that do not in fact have them.

Despite the difference just sketched between functionalism and behaviorism, functionalists and behaviorists need not be far apart in spirit.[3] Shoemaker (1975), for example, says, ''On one construal of it, functionalism in the philosophy of mind is the doctrine that mental, or psychological, terms are, in principle, eliminable in a certain way'' (pp. 306–307). Functionalists have tended to treat the mental-state terms in a functional characterization of a mental state quite differently from the input and output terms. Thus in the simplest Turing-machine version of the theory (Putnam, 1967; Block & Fodor, 1972), mental states are identified with the total Turing-machine states, which are themselves *implicitly* defined by a machine table that *explicitly* mentions inputs and outputs, described nonmentalistically.

In Lewis's version of functionalism, mental-state terms are defined by means of a modification of Ramsey's method, in a way that eliminates essential use of mental terminology from the definitions but does not eliminate input and output terminology. That is, ''pain'' is defined as synonymous with a definite description containing input and output terms but no mental terminology (see Lewis, 1972).

Furthermore, functionalism in both its machine and nonmachine versions has typically insisted that characterizations of mental states should contain descriptions of inputs and outputs in *physical* language. Armstrong (1968), for example, says,

> We may distinguish between ''physical behaviour,'' which refers to any merely physical action or passion of the body, and ''behaviour proper'' which implies relationship to the mind.... Now, if in our formula [''state of the person apt for bringing about a certain sort of behaviour''] ''behaviour'' were to mean 'behaviour proper', then we would be giving an account of mental concepts in terms of a concept that already presupposes mentality, which would be circular. So it is clear that in our formula, ''behaviour'' must mean ''physical behaviour.'' (p. 84)

Therefore, functionalism can be said to ''tack down'' mental states only at the periphery—i.e., through physical, or at least nonmental, specification of inputs and outputs. One major thesis of this article is that, because of this feature, functionalism

fails to avoid the sort of problem for which it rightly condemns behaviorism. Functionalism, too, is guilty of liberalism, for much the same reasons as behaviorism. Unlike behaviorism, however, functionalism can naturally be altered to avoid liberalism—but only at the cost of falling into an equally ignominious failing.

The failing I speak of is the one that functionalism shows *physicalism* to be guilty of. By ''physicalism,'' I mean the doctrine that pain, for example, is identical to a physical (or physiological) state.[4] As many philosophers have argued (notably Fodor, 1965, and Putnam, 1966; see also Block & Fodor, 1972), if functionalism is true, physicalism is probably false. The point is at its clearest with regard to Turing-machine versions of functionalism. Any given abstract Turing machine can be realized by a wide variety of physical devices; indeed, it is plausible that, given any putative correspondence between a Turing-machine state and a configurational physical (or physiological) state, there will be a possible realization of the Turing machine that will provide a counterexample to that correspondence. (See Kalke, 1969; Gendron, 1971; Mucciolo, 1974, for unconvincing arguments to the contrary; see also Kim, 1972.) Therefore, if pain is a functional state, it cannot, for example, be a brain state, because creatures without brains can realize the same Turing machine as creatures with brains.

I must emphasize that the functionalist argument against physicalism does not appeal merely to the fact that one abstract Turing machine can be realized by systems of different *material composition* (wood, metal, glass, etc.). To argue this way would be like arguing that temperature cannot be a microphysical magnitude because the same temperature can be had by objects with *different* microphysical structures (Kim, 1972). Objects with different microphysical structures, e.g., objects made of wood, metal, glass, etc., can have many interesting microphysical properties in common, such as molecular kinetic energy of the same average value. Rather, the functionalist argument against physicalism is that it is difficult to see how there *could be* a nontrivial first-order (see note 4) physical property in common to all and only the possible physical realizations of a given Turing-machine state. Try to think of a remotely plausible candidate! At the very least, the onus is on those who think such physical properties are conceivable to show us how to conceive of one.

One way of expressing this point is that, according to functionalism, physicalism is a *chauvinist* theory: it withholds mental properties from systems that in fact have them. In saying mental states are brain states, for example, physicalists unfairly exclude those poor brainless creatures who nonetheless have minds.

A second major point of this paper is that the very argument which functionalism uses to condemn physicalism can be applied equally well against functionalism; indeed, any version of functionalism that avoids liberalism falls, like physicalism, into chauvinism.

This article has three parts. The first argues that functionalism is guilty of liberalism, the second that one way of modifying functionalism to avoid liberalism is to tie it

more closely to empirical psychology, and the third that no version of functionalism can avoid both liberalism and chauvinism.

1.1 More about What Functionalism Is

One way of providing some order to the bewildering variety of functionalist theories is to distinguish between those that are couched in terms of a Turing machine and those that are not.

A Turing-machine table lists a finite set of machine-table states, $S_1 \ldots S_n$; inputs, $I_1 \ldots I_m$; and outputs, $O_1 \ldots O_p$. The table specifies a set of conditionals of the form: if the machine is in state S_i and receives input I_j, it emits output O_k and goes into state S_l. That is, given any state and input, the table specifies an output and a next state. Any system with a set of inputs, outputs, and states related in the way specified by the table is described by the table and is a realization of the abstract automaton specified by the table.

To have the power for computing any recursive function, a Turing machine must be able to control its input in certain ways. In standard formulations, the output of a Turing machine is regarded as having two components. It prints a symbol on a tape, then moves the tape, thus bringing a new symbol into the view of the input reader. For the Turing machine to have full power, the tape must be infinite in at least one direction and movable in both directions. If the machine has no control over the tape, it is a ''finite transducer,'' a rather limited Turing machine. Finite transducers need not be regarded as having tape at all. Those who believe that machine functionalism is true must suppose that just what power automaton we are is a substantive empirical question. If we are ''full power'' Turing machines, the environment must constitute part of the tape.

One very simple version of machine functionalism (Block and Fodor, 1972) states that each system having mental states is described by at least one Turing-machine table of a specifiable sort and that each type of mental state of the system is identical to one of the machine-table states. Consider, for example, the Turing machine described in the table (cf. Nelson, 1975):

	S_1	S_2
nickel input	Emit no output Go to S_2	Emit a Coke Go to S_1
dime input	Emit a Coke Stay in S_1	Emit a Coke & a nickel Go to S_1

One can get a crude picture of the simple version of machine functionalism by considering the claim that S_1 = dime-desire, and S_2 = nickel-desire. Of course, no functionalist would claim that a Coke machine desires anything. Rather, the simple version of machine functionalism described above makes an analogous claim with respect to a much more complex hypothetical machine table. Notice that machine functionalism specifies inputs and outputs explicitly, internal states implicitly (Putnam [1967, p. 434] says: ''The S_i, to repeat, are specified only *implicitly* by the description, i.e., specified *only* by the set of transition probabilities given in the machine table''). To be described by this machine table, a device must accept nickels and dimes as inputs and dispense nickels and Cokes as outputs. But the states S_1 and S_2 can have virtually any natures (even nonphysical natures), so long as those natures connect the states to each other and to the inputs and outputs specified in the machine table. All we are told about S_1 and S_2 are these relations; thus machine functionalism can be said to reduce mentality to input-output structures. This example should suggest the force of the functionalist argument against physicalism. Try to think of a first-order (see note 4) physical property that can be shared by all (and only) realizations of this machine table!

One can also categorize functionalists in terms of whether they regard functional identities as part of a priori psychology or empirical psychology. The a priori functionalists (e.g., Smart, Armstrong, Lewis, Shoemaker) are the heirs of the logical behaviorists. They tend to regard functional analyses as analyses of the meanings of mental terms, whereas the empirical functionalists (e.g., Fodor, Putnam, Harman) regard functional analyses as substantive scientific hypotheses. In what follows, I shall refer to the former view as ''Functionalism'' and the latter as ''Psychofunctionalism.'' (I shall use 'functionalism' with a lowercase ''f'' as neutral between Functionalism and Psychofunctionalism. When distinguishing between Functionalism and Psychofunctionalism, I shall always use capitals.)

Functionalism and Psychofunctionalism and the difference between them can be made clearer in terms of the notion of the Ramsey sentence of a psychological theory. Mental-state terms that appear in a psychological theory can be defined in various ways by means of the Ramsey sentence of the theory. All functional-state identity theories can be understood as defining a set of functional states by means of the Ramsey sentence of a psychological theory—with one functional state corresponding to each mental state. The functional state corresponding to pain will be called the ''Ramsey functional correlate'' of pain, with respect to the psychological theory. In terms of the notion of a Ramsey functional correlate with respect to a theory, the distinction between Functionalism and Psychofunctionalism can be defined as follows: Functionalism identifies mental state S with S's Ramsey functional correlate with respect to a *common-sense* psychological theory; Psychofunctionalism identifies S with S's Ramsey functional correlate with respect to a *scientific* psychological theory.

This difference between Functionalism and Psychofunctionalism gives rise to a difference in specifying inputs and outputs. Functionalists are restricted to specification of inputs and outputs that are plausibly part of commonsense knowledge; Psychofunctionalists are under no such restriction. Although both groups insist on physical—or at least nonmental—specification of inputs and outputs, Functionalists require externally observable classifications (e.g., inputs characterized in terms of objects present in the vicinity of the organism, outputs in terms of movements of body parts). Psychofunctionalists, on the other hand, have the option to specify inputs and outputs in terms of internal parameters, e.g., signals in input and output neurons.

Let T be a psychological theory of either common sense or scientific psychology. T may contain generalizations of the form: anyone who is in state w and receives input x emits output y, and goes into state z. Let us write T as

$T(S_1 \ldots S_n, I_1 \ldots I_k, O_1 \ldots O_m)$

where the S's are mental states, the I's are inputs, and the O's are outputs. The S's are to be understood as mental-state *constants*, not variables, e.g., ''pain,'' and likewise for the I's and O's. Thus, one could also write T as

T(pain . . . , light of 400 nanometers entering left eye . . . , left big toe moves 1 centimeter left . . .)

To get the Ramsey sentence of T, replace the mental-state terms—*but not the input and output terms*—by variables, and prefix an existential quantifier for each variable:

$\exists F_1 \ldots \exists F_n T(F_1 \ldots F_n, I_1 \ldots I_k, O_1 \ldots O_m)$

If F_{17} is the variable that replaced the word ''pain'' when the Ramsey sentence was formed, we can define pain as follows in terms of the Ramsey sentence:

x is in pain $\langle\!\!-\!\!\rangle$ $\exists F_1 \ldots \exists F_n T[(F_1 \ldots F_n, I_1 \ldots I_k, O_1 \ldots O_m)$ & x has $F_{17}]$

The Ramsey functional correlate of pain is the property expressed by the predicate on the right-hand side of this biconditional. Notice that this predicate contains input and output constants, but no mental constants, since the mental constants were replaced by variables. The Ramsey functional correlate for pain is defined in terms of inputs and outputs, but not in mental terms.

For example, let T be the theory that pain is caused by skin damage and causes worry and the emission of ''Ouch,'' and worry, in turn, causes brow wrinkling. Then the Ramsey definition would be:

x is in pain $\langle\!\!-\!\!\rangle$ There are two states (properties), the first of which is caused by skin damage and causes both the emission of ''Ouch'' and the second state, and the second state causes brow wrinkling, and x is in the first state.

The Ramsey functional correlate of pain with respect to this ''theory'' is the property of being in a state that is caused by skin damage and causes the emission of ''ouch'' and another state that in turn causes brow wrinkling. (Note that the words ''pain'' and ''worry'' have been replaced by variables, but the input and output terms remain.)

The Ramsey functional correlate of a state S is a state that has much in common with S. Specifically, S and its Ramsey functional correlate share the structural properties specified by the theory T. But there are two reasons why it is natural to suppose that S and its Ramsey functional correlate will be distinct. First, the Ramsey functional correlate of S with respect to T can ''include'' at most those aspects of S that are captured by T; any aspects not captured by T will be left out. Second, the Ramsey functional correlate may even leave out some of what T does capture, for the Ramsey definition does not contain the ''theoretical'' vocabulary of T. The example theory of the last paragraph is true only of pain-feeling organisms—but trivially, by virtue of its use of the word ''pain.'' However, the predicate that expresses T's Ramsey functional correlate does not contain this word (since it was replaced by a variable), and so can be true of things that do not feel pain. It would be easy to make a simple machine that has some artificial skin, a brow, a tape-recorded ''ouch,'' and two states that satisfy the mentioned causal relations, but no pain.

The bold hypothesis of functionalism is that for *some* psychological theory, this natural supposition that a state and its Ramsey functional correlate are distinct is false. Functionalism says that there is a theory such that pain, for example, *is* its Ramsey functional correlate with respect to that theory.

One final preliminary point: I have given the misleading impression that functionalism identifies *all* mental states with functional states. Such a version of functionalism is obviously far too strong. Let X be a newly created cell-for-cell duplicate of you (which, of course, is functionally equivalent to you). Perhaps you remember being bar-mitzvahed. But X does not remember being bar-mitzvahed, since X never was bar-mitzvahed. Indeed, something can be functionally equivalent to you but fail to know what you know, or [verb], what you [verb], for a wide variety of ''success'' verbs. Worse still, if Putnam (1975b) is right in saying that ''meanings are not in the head,'' systems functionally equivalent to you may, for similar reasons, fail to have many of your other propositional attitudes. Suppose you believe water is wet. According to plausible arguments advanced by Putnam and Kripke, a condition for the possibility of your believing water is wet is a certain kind of causal connection between you and water. Your ''twin'' on Twin Earth, who is connected in a similar way to XYZ rather than H_2O, would not believe water is wet.

If functionalism is to be defended, it must be construed as applying only to a subclass of mental states, those ''narrow'' mental states such that truth conditions for their application are in some sense ''within the person.'' But even assuming that a notion of narrowness of psychological state can be satisfactorily formulated, the interest

of functionalism may be diminished by this restriction. I mention this problem only to set it aside.

I shall take functionalism to be a doctrine about all ''narrow'' mental states.

1.2 Homunculi-Headed Robots

In this section I shall describe a class of devices that are *prima facie* embarrassments for all versions of functionalism in that they indicate functionalism is guilty of liberalism—classifying systems that lack mentality as having mentality.

Consider the simple version of machine functionalism already described. It says that each system having mental states is described by at least one Turing-machine table of a certain kind, and each mental state of the system is identical to one of the machine-table states specified by the machine table. I shall consider inputs and outputs to be specified by descriptions of neural impulses in sense organs and motor-output neurons. This assumption should not be regarded as restricting what will be said to Psychofunctionalism rather than Functionalism. As already mentioned, every version of functionalism assumes *some* specificiation of inputs and outputs. A Functionalist specification would do as well for the purposes of what follows.

Imagine a body externally like a human body, say yours, but internally quite different. The neurons from sensory organs are connected to a bank of lights in a hollow cavity in the head. A set of buttons connects to the motor-output neurons. Inside the cavity resides a group of little men. Each has a very simple task: to implement a ''square'' of an adequate machine table that describes you. On one wall is a bulletin board on which is posted a state card, i.e., a card that bears a symbol designating one of the states specified in the machine table. Here is what the little men do: Suppose the posted card has a 'G' on it. This alerts the little men who implement G squares—'G-men' they call themselves. Suppose the light representing input I_{17} goes on. One of the G-men has the following as his sole task: when the card reads 'G' and the I_{17} light goes on, he presses output button O_{191} and changes the state card to 'M'. This G-man is called upon to exercise his task only rarely. In spite of the low level of intelligence required of each little man, the system as a whole manages to simulate you because the functional organization they have been trained to realize is yours. A Turing machine can be represented as a finite set of quadruples (or quintuples, if the output is divided into two parts): current state, current input; next state, next output. Each little man has the task corresponding to a single quadruple. Through the efforts of the little men, the system realizes the same (reasonably adequate) machine table as you do and is thus functionally equivalent to you.[6]

I shall describe a version of the homunculi-headed simulation, which has more chance of being nomologically possible. How many homunculi are required? Perhaps a billion are enough.

Suppose we convert the government of China to functionalism, and we convince its officials to realize a human mind for an hour. We provide each of the billion people in China (I chose China because it has a billion inhabitants) with a specially designed two-way radio that connects them in the appropriate way to other persons and to the artificial body mentioned in the previous example. We replace each of the little men with a citizen of China plus his radio. Instead of a bulletin board, we arrange to have letters displayed on a series of satellites placed so that they can be seen from anywhere in China.

The system of a billion people communicating with one another plus satellites plays the role of an external ''brain'' connected to the artificial body by radio. There is nothing absurd about a person being connected to his brain by radio. Perhaps the day will come when our brains will be periodically removed for cleaning and repairs. Imagine that this is done initially by treating neurons attaching the brain to the body with a chemical that allows them to stretch like rubber bands, thereby assuring that no brain-body connections are disrupted. Soon clever businessmen discover that they can attract more customers by replacing the stretched neurons with radio links so that brains can be cleaned without inconveniencing the customer by immobilizing his body.

It is not at all obvious that the China-body system is physically impossible. It could be functionally equivalent to you for a short time, say an hour.

''But,'' you may object, ''how could something be functionally equivalent to me for *an hour?* Doesn't my functional organization determine, say, how I would react to doing nothing for a week but reading the *Reader's Digest?*'' Remember that a machine table specifies a set of conditionals of the form: if the machine is in S_i and receives input I_j, it emits output O_k and goes into S_l. These conditionals are to be understood *subjunctively*. What gives a system a functional organization at a time is not just what it *does* at that time, but also the counterfactuals true of it at that time: what it *would* have done (and what its state transitions would have been) had it had a different input or been in a different state. If it is true of a system at time t that it *would* obey a given machine table no matter which of the states it is in and no matter which of the inputs it receives, then the system is described at t by the machine table (and realizes at t the abstract automaton specified by the table), even if it exists for only an instant. For the hour the Chinese system is ''on,'' it *does* have a set of inputs, outputs, and states of which such subjunctive conditionals are true. This is what makes any computer realize the abstract automaton that it realizes.

Of course, there are signals the system would respond to that you would not respond to, e.g., massive radio interference or a flood of the Yangtze River. Such events might cause a malfunction, scotching the simulation, just as a bomb in a computer can make it fail to realize the machine table it was built to realize. But just as the computer *without* the bomb *can* realize the machine table, the system consisting of the people

and artificial body can realize the machine table so long as there are no catastrophic interferences, e.g., floods, etc.

''But,'' someone may object, ''there is a difference between a bomb in a computer and a bomb in the Chinese system, for in the case of the latter (unlike the former), inputs as specified in the machine table can be the cause of the malfunction. Unusual neural activity in the sense organs of residents of Chungking Province caused by a bomb or by a flood of the Yangtze can cause the system to go haywire.''

Reply: The person who says what system he or she is talking about gets to say what signals count as inputs and outputs. I count as inputs and outputs only neural activity in the artificial body connected by radio to the people of China. Neural signals in the people of Chungking count no more as inputs to this system than input tape jammed by a saboteur between the relay contacts in the innards of a computer count as an input to the computer.

Of course, the object consisting of the people of China + the artificial body has *other* Turing-machine descriptions under which neural signals in the inhabitants of Chungking *would* count as inputs. Such a new system (i.e., the object under such a new Turing-machine description) would not be functionally equivalent to you. Likewise, any commercial computer can be redescribed in a way that allows tape jammed into its innards to count as inputs. In describing an object as a Turing machine, one draws a line between the inside and the outside. (If we count only neural impulses as inputs and outputs, we draw that line inside the body; if we count only peripheral stimulations as inputs, we draw that line at the skin.) In describing the Chinese system as a Turing machine, I have drawn the line in such a way that it satisfies a certain type of functional description—one that you *also* satisfy, and one that, according to functionalism, justifies attributions of mentality. Functionalism does not claim that every mental system has a machine table of a sort that justifies attributions of mentality with respect to *every* specification of inputs and outputs, but rather, only with respect to *some* specification.

Objection: The Chinese system would work too slowly. The kind of events and processes with which we normally have contact would pass by far too quickly for the system to detect them. Thus, we would be unable to converse with it, play bridge with it, etc.

Reply: It is hard to see why the system's time scale should matter. Is it really contradictory or nonsensical to suppose we could meet a race of intelligent beings with whom we could communicate only by devices such as time-lapse photography? When we observe these creatures, they seem almost inanimate. But when we view the time-lapse movies, we see them conversing with one another. Indeed, we find they are saying that the only way they can make any sense of us is by viewing movies greatly slowed down. To take time scale as all important seems crudely behavioristic. Further, even if the time-scale objection is right, I can elude it by retreating to the point that a

homunculi-head that works in normal time is *metaphysically* possible. Metaphysical possibility is all my argument requires. (See Kripke, 1972.)

What makes the homunculi-headed system (count the two systems as variants of a single system) just described a prima facie counterexample to (machine) functionalism is that there is prima facie doubt whether it has any mental states at all—especially whether it has what philosophers have variously called ''qualitative states,'' ''raw feels,'' or ''immediate phenomenological qualities.'' (You ask: What is it that philosophers have called qualitative states? I answer, only half in jest: As Louis Armstrong said when asked what jazz is, ''If you got to ask, you ain't never gonna get to know.'') In Nagel's terms (1974), there is a prima facie doubt whether there is anything which it is like to be the homunculi-headed system.

The force of the prima facie counterexample can be made clearer as follows: Machine functionalism says that each mental state is identical to a machine-table state. For example, a particular qualitative state, Q, is identical to a machine-table state, S_q. But if there is nothing it is like to be the homunculi-headed system, it cannot be in Q even when it is in S_q. Thus, if there is prima facie doubt about the homunculi-headed system's mentality, there is prima facie doubt that $Q = S_q$, i.e., doubt that the kind of functionalism under consideration is true.[7] Call this argument the Absent Qualia Argument.

So there is prima facie doubt that machine functionalism is true. So what? After all, prima facie doubt is only prima facie. Indeed, appeals to intuition of this sort are notoriously fallible. I shall not rest on this appeal to intuition. Rather, I shall argue that the intuition that the homunculi-headed simulation described above lacks mentality (or at least qualia) has at least in part a rational basis, and that this rational basis provides a good reason for doubting that Functionalism (and to a lesser degree Psychofunctionalism) is true. I shall consider this line of argument in Section 1.5.

1.3 What If I Turned Out to Have Little Men in My Head?

Before I go any further, I shall briefly discuss a difficulty for my claim that there is prima facie doubt about the qualia of homunculi-headed realizations of human functional organization. It might be objected, ''What if *you* turned out to be one?'' Let us suppose that, to my surprise, X-rays reveal that inside my head are thousands of tiny, trained fleas, each of which has been taught (perhaps by a joint subcommittee of the American Philosophical Association and the American Psychological Association empowered to investigate absent qualia) to implement a square in the appropriate machine table.

Now there is a crucial issue relevant to this difficulty which philosophers are far from agreeing on (and about which I confess I cannot make up my mind): Do I know on the basis of my ''privileged access'' that I do not have utterly absent qualia, no matter what

turns out to be inside my head? Do I know there is something it is like to be me, even if I am a flea-head? Fortunately, my vacillation on this issue is of no consequence, for either answer is compatible with the Absent Qualia Argument's assumption that there is doubt about the qualia of homunculi-headed folks.

Suppose the answer is no. It is not the case that I know there is something it is like to be me even if I am a flea-head. Then I should admit that my qualia would be in (prima facie) doubt if (God forbid) I turned out to have fleas in my head. Likewise for the qualia of all the other homunculi-headed folk. So far, so good.

Suppose, on the other hand, that my privileged access does give me infallible knowledge that I have qualia. No matter what turns out to be inside my head, my states have qualitative content. There is something it is like to be me. Then if I turn out to have fleas in my head, at least one homunculi-head turns out to have qualia. But this would not challenge my claim that the qualia of homunculi-infested simulations is in doubt. Since I do, in fact, have qualia, supposing I have fleas inside my head is supposing someone with fleas inside his head has qualia. But this supposition that a homunculi-head has qualia is just the sort of supposition my position doubts. Using such an example to argue against my position is like twitting a man who doubts there is a God by asking what he would say if he turned out to *be* God. Both arguments against the doubter beg the question against the doubter by hypothesizing a situation which the doubter admits is logically possible, but doubts is *actual*. A doubt that there is a God entails a doubt that I am God. Similarly, (given that I do have qualia) a doubt that flea-heads have qualia entails a doubt that I am a flea-head.

1.4 Putnam's Proposal

One way functionalists can try to deal with the problem posed by the homunculi-headed counterexamples is by the ad hoc device of stipulating them away. For example, a functionalist might stipulate that two systems cannot be functionally equivalent if one contains parts with functional organizations characteristic of sentient beings and the other does not. In his article hypothesizing that pain is a functional state, Putnam stipulated that ''no organism capable of feeling pain possesses a decomposition into parts which separately possess Descriptions'' (as the sort of Turing machine which can be in the functional state Putnam identifies with pain). The purpose of this condition is ''to rule out such 'organisms' (if they count as such) as swarms of bees as single pain feelers'' (Putnam, 1967, pp. 434–435).

One way of filling out Putnam's requirement would be: a pain-feeling organism cannot possess a decomposition into parts *all* of which have a functional organization characteristic of sentient beings. But this would not rule out my homunculi-headed example, since it has nonsentient parts, such as the mechanical body and sense organs. It will not do to go to the opposite extreme and require that *no* proper parts be sentient.

Otherwise pregnant women and people with sentient parasites will fail to count as pain-feeling organisms. What seems to be important to examples like the homunculi-headed simulation I have described is that the sentient beings *play a crucial role* in giving the thing its functional organization. This suggests a version of Putnam's proposal which requires that a pain-feeling organism has a certain functional organization and that it has no parts which (1) themselves possess that sort of functional organization and also (2) play a crucial role in giving the whole system its functional organization.

Although this proposal involves the vague notion ''crucial role,'' it is precise enough for us to see it will not do. Suppose there is a part of the universe that contains matter quite different from ours, matter that is infinitely divisible. In this part of the universe, there are intelligent creatures of many sizes, even humanlike creatures much smaller than our elementary particles. In an intergalactic expedition, these people discover the existence of our type of matter. For reasons known only to them, they decide to devote the next few hundred years to creating out of *their* matter substances with the chemical and physical characteristics (except at the subelementary particle level) of *our* elements. They build hordes of space ships of different varieties about the sizes of our electrons, protons, and other elementary particles, and fly the ships in such a way as to mimic the behavior of these elementary particles. The ships also contain generators to produce the type of radiation elementary particles give off. Each ship has a staff of experts on the nature of our elementary particles. They do this so as to produce huge (by our standards) masses of substances with the chemical and physical characteristics of oxygen, carbon, etc. Shortly after they accomplish this, you go off on an expedition to that part of the universe, and discover the ''oxygen,'' ''carbon,'' etc. Unaware of its real nature, you set up a colony, using these ''elements'' to grow plants for food, provide ''air'' to breathe, etc. Since one's molecules are constantly being exchanged with the environment, you and other colonizers come (in a period of a few years) to be composed mainly of the ''matter'' made of the tiny people in space ships. Would you be any less capable of feeling pain, thinking, etc. just because the matter of which you are composed contains (and depends on for its characteristics) beings who themselves have a functional organization characteristic of sentient creatures? I think not. The basic electrochemical mechanisms by which the synapse operates are now fairly well understood. As far as is known, changes that do not affect these electrochemical mechanisms do not affect the operation of the brain, and do not affect mentality. The electrochemical mechanisms in your synapses would be unaffected by the change in your matter.[8]

It is interesting to compare the elementary-particle-people example with the homunculi-headed examples the chapter started with. A natural first guess about the source of our intuition that the initially described homunculi-headed simulations lack mentality is that they have *too much* internal mental structure. The little men may be sometimes bored, sometimes excited. We may even imagine that they deliberate about

the best way to realize the given functional organization and make changes intended to give them more leisure time. But the example of the elementary-particle people just described suggests this first guess is wrong. What seems important is *how* the mentality of the parts contributes to the functioning of the whole.

There is one very noticeable difference between the elementary-particle-people example and the earlier homunculus examples. In the former, the change in you as you become homunculus-infested is not one that makes any difference to your psychological processing (i.e., information processing) or neurological processing but only to your microphysics. No techniques proper to human psychology or neurophysiology would reveal any difference in you. However, the homunculi-headed simulations described in the beginning of the chapter are not things to which neurophysiological theories true of us apply, and *if they are construed as Functional* (rather than Psychofunctional) simulations, they need not be things to which psychological (information-processing) theories true of us apply. This difference suggests that our intuitions are in part controlled by the not unreasonable view that our mental states depend on our having the psychology and/or neurophysiology we have. So something that differs markedly from us in both regards (recall that it is a Functional rather than Psychofunctional simulation) should not be assumed to have mentality just on the ground that it has been designed to be Functionally equivalent to us.[9]

1.5 Is the Prima Facie Doubt Merely Prima Facie?

The Absent Qualia Argument rested on an appeal to the intuition that the homunculi-headed simulations lacked mentality, or at least qualia. I said that this intuition gave rise to prima facie doubt that functionalism is true. But intuitions unsupported by principled argument are hardly to be considered bedrock. Indeed, intuitions incompatible with well-supported theory (e.g., the pre-Copernican intuition that the earth does not move) thankfully soon disappear. Even fields like linguistics whose data consist mainly in intuitions often reject such intuitions as that the following sentences are ungrammatical (on theoretical grounds):

The horse raced past the barn fell.
The boy the girl the cat bit scratched died.

These sentences are in fact grammatical though hard to process.[10]

Appeal to intuitions when judging possession of mentality, however, is *especially* suspicious. *No* physical mechanism seems very intuitively plausible as a seat of qualia, least of all a *brain*. Is a hunk of quivering gray stuff more intuitively appropriate as a seat of qualia than a covey of little men? If not, perhaps there is a prima facie doubt about the qualia of brain-headed systems too?

However, there is a very important difference between brain-headed and homunculi-headed systems. Since we know that *we are brain-headed systems*, and that *we* have qualia, we know that brain-headed systems can have qualia. So even though we have no theory of qualia which explains how this is *possible*, we have overwhelming reason to disregard whatever prima facie doubt there is about the qualia of brain-headed systems. Of course, this makes my argument partly *empirical*—it depends on knowledge of what makes us tick. But since this is knowledge we in fact possess, dependence on this knowledge should not be regarded as a defect.[11]

There is another difference between us meat-heads and the homunculi-heads: they are systems designed to mimic us, but we are not designed to mimic anything (here I rely on another empirical fact). This fact forestalls any attempt to argue on the basis of an inference to the best explanation for the qualia of homunculi-heads. The best explanation of the homunculi-heads' screams and winces is not their pains, but that they were designed to mimic our screams and winces.

Some people seem to feel that the complex and subtle behavior of the homunculi-heads (behavior just as complex and subtle—even as ''sensitive'' to features of the environment, human and nonhuman, as your behavior) is itself sufficient reason to disregard the prima facie doubt that homunculi-heads have qualia. But this is just crude behaviorism.

My case against functionalism depends on the following principle: if a doctrine has an absurd conclusion that there is no independent reason to believe, and if there is no way of explaining away the absurdity or showing it to be misleading or irrelevant, and if there is no good reason to believe the doctrine that leads to the absurdity in the first place, then don't accept the doctrine. I claim that there is no independent reason to believe in the mentality of the homunculi-head, and I know of no way of explaining away the absurdity of the conclusion that it has mentality (though of course, my argument is vulnerable to the introduction of such an explanation). The issue, then, is whether there is any good reason to believe Functionalism. One argument for Functionalism is that it is the best solution available to the mind-body problem. I think this is a bad form of argument, but since I also think that Psychofunctionalism is preferable to Functionalism (for reasons to be mentioned below), I will postpone consideration of this form of argument to the discussion of Psychofunctionalism.

The only other argument for Functionalism that I know of is that Functional identities can be shown to be true on the basis of analyses of the meanings of mental terminology. According to this argument, Functional identities are to be justified in the way one might try to justify the claim that the state of being a bachelor is identical to the state of being an unmarried man. A similar argument appeals to commonsense platitudes about mental states instead of truths of meaning. Lewis says that Functional characterizations of mental states are in the province of ''common sense psychology—

folk science, rather than professional science'' (Lewis, 1972, p. 250). (See also Shoemaker, 1975, and Armstrong, 1968. Armstrong equivocates on the analyticity issue. See Armstrong, 1968, pp. 84–85, and p. 90.) And he goes on to insist that Functional characterizations should ''include only platitudes which are common knowledge among us—everyone knows them, everyone knows that everyone else knows them, and so on'' (Lewis, 1972, p. 256). I shall talk mainly about the ''platitude'' version of the argument. The analyticity version is vulnerable to essentially the same considerations, as well as Quinean doubts about analyticity.

I am willing to concede, for the sake of argument, that it is possible to define any given mental-state term in terms of platitudes concerning other mental-state terms, input terms, and output terms. But this does not commit me to the type of definition of mental terms in which all mental terminology has been eliminated via Ramsification or some other device. It is simply a fallacy to suppose that if each mental term is definable in terms of the others (plus inputs and outputs), then each mental term is definable nonmentalistically. To see this, consider the example given earlier. Indeed, let's simplify matters by ignoring the inputs and outputs. Let's define pain as the cause of worry, and worry as the effect of pain. Even a person so benighted as to accept this, need not accept a definition of pain as *the cause of something*, or a definition of worry as *the effect of something*. Lewis claims that it is analytic that pain is the occupant of a certain causal role. Even if he is right about a causal role, *specified in part mentalistically*, one cannot conclude that it is analytic that pain is the occupant of any causal role, nonmentalistically specified.

I do not see any decent argument for Functionalism based on platitudes or analyticity. Further, the conception of Functionalism as based on platitudes leads to trouble with cases that platitudes have nothing to say about.

Recall the example of brains being removed for cleaning and rejuvenation, the connections between one's brain and one's body being maintained by radio while one goes about one's business. The process takes a few days, and when it is completed, the brain is reinserted in the body. Occasionally it may happen that a person's body is destroyed by an accident while the brain is being cleaned and rejuvenated. If hooked up to input sense organs (but not output organs) such a brain would exhibit *none* of the usual platitudinous connections between behavior and clusters of inputs and mental states. If, as seems plausible, such a brain could have almost all the same (narrow) mental states as we have (and since such a state of affairs could become typical), Functionalism is wrong.

It is instructive to compare the way Psychofunctionalism attempts to handle cases like paralysis and brains in bottles. According to Psychofunctionalism, what is to count as a system's inputs and outputs is an empirical question. Counting neural impulses as inputs and outputs would avoid the problems just sketched, since the brains in bottles and paralytics could have the right neural impulses even without bodily movements.

Objection: There could be paralysis that affects the nervous system, and thus affects the neural impulses, so the problem which arises for Functionalism arises for Psychofunctionalism as well. Reply: Nervous system diseases can actually *change mentality*, e.g., they can render victims incapable of having pain. So it might actually be true that a widespread nervous system disease that caused intermittent paralysis rendered people incapable of certain mental states.

According to plausible versions of Psychofunctionalism, the job of deciding what neural processes should count as inputs and outputs is in part a matter of deciding *what malfunctions count as changes in mentality and what malfunctions count as changes in peripheral input and output connections*. Psychofunctionalism has a resource that Functionalism does not have, since Psychofunctionalism allows us to *adjust the line we draw between the inside and the outside of the organism so as to avoid problems of the sort discussed*. All versions of Functionalism go wrong in attempting to draw this line on the basis of only common-sense knowledge; ''analyticity'' versions of Functionalism go especially wrong in attempting to draw the line a priori.

Objection: Sydney Shoemaker suggests (in correspondence) that problems having to do with paralytics, and brains in vats of the sort I mentioned, can be handled using his notion of a ''paradigmatically embodied person'' (see Shoemaker, 1976). Paradigmatic embodiment involves having functioning sensory apparatus and considerable voluntary control of bodily movements. Shoemaker's suggestion is that we start with a functional characterization of a paradigmatically embodied person, saying, inter alia, what it is for a physical state to realize a given mental state in a paradigmatically embodied person. Then, the functional characterization could be extended to nonparadigmatically embodied persons by saying that a physical structure that is not a part of a paradigmatically embodied person will count as realizing mental states, if, without changing its internal structure and the sorts of relationships that hold between its states, it could be incorporated into a larger physical system that would be the body of a paradigmatically embodied person in which the states in question played the functional roles definitive of mental states of a paradigmatically embodied person. Shoemaker suggests that a brain in a vat can be viewed from this perspective, as a limiting case of an amputee—amputation of everything but the brain. For the brain can (in principle) be incorporated into a system so as to form a paradigmatically embodied person without changing the internal structure and state relations of the brain.

Reply: Shoemaker's suggestion is very promising, but it saves functionalism only by retreating from Functionalism to Psychofunctionalism. Obviously, nothing in prescientific common-sense wisdom about mentality tells us what can or cannot be paradigmatically embodied *without changing its internal structure and state relations* (unless 'state relations' means 'Functional state relations', in which case the question is begged). Indeed, the scientific issues involved in answering this question are very similar to the scientific issues involved in the Psychofunctionalist question about the difference

between defects in or damage to input-output devices, as opposed to defects in or damage to central mechanisms. That is, the scientific task of drawing the Psychofunctionalist line between the inside and the outside of an organism are much the same as Shoemaker's task of drawing the line between what can and what cannot be paradigmatically embodied without changing its internal structure and state relations.

I shall briefly raise two additional problems for Functionalism. The first might be called the Problem of Differentiation: there are mental states that are different, but that do not differ with respect to platitudes. Consider different tastes or smells that have typical causes and effects, but whose typical causes and effects are not known or are not known to very many people. For example, tannin in wine produces a particular taste immediately recognizable to wine drinkers. As far as I know, there is no standard name or description (except ''tannic'') associated with this taste. The causal antecedents and consequents of this taste are not widely known, there are no platitudes about its typical causes and effects. Moreover, there are sensations that not only have no standard names but whose causes and effects are not yet well understood by anyone. Let A and B be two such (different) sensations. Neither platitudes nor truths of meaning can distinguish between A and B. Since the Functional description of a mental state is determined by the platitudes true of that state, and since A and B do not differ with respect to platitudes, Functionalists would be committed to identifying A and B with the same Functional state, and thus they would be committed to the claim that A = B, which is ex hypothesi false.

A second difficulty for Functionalism is that platitudes are often wrong. Let us call this problem the Problem of Truth. Lewis suggests, by way of dealing with this problem, that we specify the causal relations among mental states, inputs and outputs, not by means of the conjunction of all the platitudes, but rather by ''a cluster of them—a disjunction of conjunctions of *most* of them (that way it will not matter if a few are wrong.)'' This move may exacerbate the problem of Differentiation, however, since there may be pairs of different mental states that are alike with respect to *most* platitudes.

2.0 Psychofunctionalism

In criticizing Functionalism, I appealed to the following principle: if a doctrine has an absurd conclusion that there is no independent reason to believe, and if there is no way of explaining away the absurdity or showing it to be misleading or irrelevant, and if there is no good reason to believe the doctrine that leads to the absurdity in the first place, then do not accept the doctrine. I said that there was no independent reason to believe that the homunculi-headed Functional simulation has any mental states. However, there *is* an independent reason to believe that the homunculi-headed *Psycho*functional simulation has mental states, namely, that a Psychofunctional simu-

lation of you would be Psychofunctionally equivalent to you, so any psychological theory true of you would be true of it too. What better reason could there be to attribute to it whatever mental states are in the domain of psychology?

Even if this point shows that any Psychofunctional simulation of you shares your *non*qualitative mental states. I will argue that there is nonetheless some doubt that it shares your qualitative mental states.

Here is one argument for Psychofunctionalism that is implicit in the literature. It is the business of branches of science to tell us the nature of things in the branches' domains. Mental states are in the domain of psychology, and, hence, it is the business of psychology to tell us what mental states are. Psychological theory can be expected to characterize mental states in terms of the causal relations among mental states, and other mental entities, and among mental entities, inputs, and outputs. But these very causal relations are the ones which constitute the Psychofunctional states that Psychofunctionalism identifies with mental states. So Psychofunctionalism is just the result of applying a plausible conception of science to mentality; Psychofunctionalism is just the doctrine that mental states are the ''psychological states'' it is the business of psychology to characterize.

That something is seriously amiss with this form of argument can be seen by noting that it would be fallacious if applied to other branches of science.

Consider the analogue of Psychofunctionalism for physics. It says that protonhood, for example, is the property of having certain lawlike relations to certain other physical properties. With respect to current physical theory, protonhood would be identified with a property expressible in terms of the Ramsey sentence of current physical theory (in the manner described above). Now there is an obvious problem with this claim about what it is to be a proton. Namely, this physico-functionalist approach would identify being an anti-proton *with the very same property*. According to current physical theory, protons and antiprotons are ''dual'' entities: one cannot distinguish the variable which replaced 'protonhood' from the variable that replaced 'antiprotonhood' (in any nontrivial way) in the Ramsey sentence of current physical theory. Yet protons and anti-protons are different types of particles; it is a law of physics that particles annihilate their antiparticles; thus, protons annihilate antiprotons, even though protons get along fine with other protons.[12]

Suppose someone were to argue that ''protonhood = its Ramsey functional correlate with respect to current physical theory'' is our best hypothesis as to the nature of protonhood, on the ground that this identification amounts to an application of the doctrine that it is the business of branches of science to tell us the nature of things in their domains. The person would be arguing fallaciously. So why should we suppose that this form of argument is any less fallacious when applied to psychology?

In the preceding few paragraphs I may have given the impression that the analogue of Psychofunctionalism in physics can be used to cast doubt on Psychofunctionalism

itself. But there are two important disanalogies between Psychofunctionalism and its physics analogue. First, according to Psychofunctionalism, there is a theoretically principled distinction between, on one hand, the inputs and outputs described explicitly in the Ramsey sentence, and, on the other hand, the internal states and other psychological entities whose names are replaced by variables. But there is no analogous distinction with respect to other branches of science. An observational/theoretical distinction would be analogous if it could be made out, but difficulties in drawing such a distinction are notorious.

Second, and more important, Psychofunctionalism simply need not be regarded as a special case of any general doctrine about the nature of the entities scientific theories are about. Psychofunctionalists can reasonably hold that only *mental* entities—or perhaps only states, events, and their ilk, as opposed to substances like protons—are ''constituted'' by their causal relations. Of course, if Psychofunctionalists take such a view, they protect Psychofunctionalism from the proton problem at the cost of abandoning the argument that Psychofunctionalism is just the result of applying a plausible conception of science to mentality.

Another argument for Psychofunctionalism (or, less plausibly, for Functionalism) which can be abstracted from the literature is an ''inference to the best explanation'' argument: ''What *else* could mental states be if not Psychofunctional states?'' For example, Putnam (1967) hypothesizes that (Psycho)functionalism is true and then argues persuasively that (Psycho)functionalism is a *better* hypothesis than behaviorism or materialism.

But this is a very dubious use of ''inference to the best explanation.'' For what guarantee do we have that *there is* an answer to the question ''What are mental states?'' of the sort behaviorists, materialists, and functionalists have wanted? Moreover, inference to the best explanation cannot be applied when none of the available explanations is any good. In order for inference to the best explanation to be applicable, two conditions have to be satisfied: we must have reason to believe an explanation is *possible*, and at least one of the available explanations must be *minimally adequate*. Imagine someone arguing for one of the proposed solutions to Newcomb's Problem on the ground that despite its fatal flaw it is the best of the proposed solutions. That would be a joke. But is the argument for functionalism any better? Behaviorism, materialism, and functionalism are not theories of mentality in the way Mendel's theory is a theory of heredity. Behaviorism, materialism, and functionalism (and dualism as well) are attempts to solve a problem: the mind-body problem. Of course, this is a problem which can hardly be guaranteed to have a solution. Further, each of the proposed solutions to the mind-body problem has serious difficulties, difficulties I for one am inclined to regard as fatal.

Why is functionalism so widely accepted, given the dearth of good arguments for it, implicit or explicit? In my view, what has happened is that functionalist doctrines

were offered initially as hypotheses. But with the passage of time, plausible-sounding hypotheses with useful features can come to be treated as established facts, even if no good arguments have ever been offered for them.

2.1 Are Qualia Psychofunctional States?

I began this chapter by describing a homunculi-headed device and claiming there is prima facie doubt about whether it has any mental states at all, especially whether it has qualitative mental states like pains, itches, and sensations of red. The special doubt about qualia can perhaps be explicated by thinking about *inverted* qualia rather than *absent* qualia. It makes sense, or seems to make sense, to suppose that objects we both call green look to me the way objects we both call red look to you. It seems that we could be functionally equivalent even though the sensation fire hydrants evoke in you is qualitatively the same as the sensation grass evokes in me. Imagine an inverting lens which when placed in the eye of a subject results in exclamations like ''Red things now look the way green things used to look, and vice versa.'' Imagine further, a pair of identical twins one of whom has the lenses inserted at birth. The twins grow up normally, and at age 21 are functionally equivalent. This situation offers at least some evidence that each's spectrum is inverted relative to the other's. (See Shoemaker, 1975, note 17, for a convincing description of intrapersonal spectrum inversion.) However, it is very hard to see how to make sense of the analogue of spectrum inversion with respect to nonqualitative states. Imagine a pair of persons one of whom believes that p is true and that q is false, while the other believes that q is true and that p is false. Could these persons be functionally equivalent? It is hard to see how they could.[13] Indeed, it is hard to see how two persons could have only this difference in beliefs and yet there be no possible circumstance in which this belief difference would reveal itself in different behavior. Beliefs seem to be supervenient on functional organization in a way that qualia are not.

There is another reason to firmly distinguish between qualitative and nonqualitative mental states in talking about functionalist theories: Psychofunctionalism avoids Functionalism's problems with nonqualitative states, e.g., propositional attitudes like beliefs and desires. But Psychofunctionalism may be no more able to handle qualitative states than is Functionalism. The reason is that qualia may well not be in the domain of psychology.

To see this, let us try to imagine what a homunculi-headed realization of human psychology would be like. Current psychological theorizing seems directed toward the description of information-flow relations among psychological mechanisms. The aim seems to be to decompose such mechanisms into psychologically primitive mechanisms, ''black boxes'' whose internal structure is in the domain of physiology rather than in the domain of psychology. (See Fodor, 1968b, Dennett, 1975, and

Cummins, 1975; interesting objections are raised in Nagel, 1969.) For example, a near-primitive mechanism might be one that matches two items in a representational system and determines if they are tokens of the same type. Or the primitive mechanisms might be like those in a digital computer, e.g., they might be (a) *add 1 to a given register*, and (b) *subtract 1 from a given register, or if the register contains 0, go to the nth (indicated) instruction*. (These operations can be combined to accomplish any digital computer operation; see Minsky, 1967, p. 206.) Consider a computer whose machine-language code contains only two instructions corresponding to (a) and (b). If you ask how it multiplies or solves differential equations or makes up payrolls, you can be answered by being shown a program couched in terms of the two machine-language instructions. But if you ask how it adds 1 to a given register, the appropriate answer is given by a wiring diagram, not a program. The machine is hard-wired to add 1. When the instruction corresponding to (a) appears in a certain register, the contents of another register ''automatically'' change in a certain way. The computational structure of a computer is determined by a set of primitive operations and the ways nonprimitive operations are built up from them. Thus it does not matter to the computational structure of the computer whether the primitive mechanisms are realized by tube circuits, transistor circuits, or relays. Likewise, it does not matter to the psychology of a mental system whether its primitive mechanisms are realized by one or another neurological mechanism. Call a system a ''realization of human psychology'' if every psychological theory true of us is true of it. Consider a realization of human psychology whose primitive psychological operations are accomplished by little men, in the manner of the homunculi-headed simulations discussed. So, perhaps one little man produces items from a list, one by one, another compares these items with other representations to determine whether they match, etc.

Now there is good reason for supposing this system has some mental states. Propositional attitudes are an example. Perhaps psychological theory will identify remembering that P with having ''stored'' a sentencelike object which expresses the proposition that P (Fodor, 1975). Then if one of the little men has put a certain sentencelike object in ''storage,'' we may have reason for regarding the system as remembering that P. But unless having qualia is just a matter of having certain information processing (at best a controversial proposal—see later discussion), there is no such theoretical reason for regarding the system as having qualia. In short, there is perhaps as much doubt about the qualia of this homunculi-headed system as there was about the qualia of the homunculi-headed Functional simulation discussed early in the chapter.

But the system we are discussing is ex hypothesi something of which any true psychological theory is true. *So any doubt that it has qualia is a doubt that qualia are in the domain of psychology*.

It may be objected: ''The kind of psychology you have in mind is *cognitive* psychology, i.e., psychology of thought processes; and it is no wonder that qualia are not in

the domain of *cognitive* psychology!'' But I *do not* have cognitive psychology in mind, and if it sounds that way, this is easily explained: nothing we know about the psychological processes underlying our conscious mental life has anything to do with qualia. What passes for the ''psychology'' of sensation or pain, for example, is (a) physiology, (b) psychophysics (i.e., study of the mathematical functions relating stimulus variables and sensation variables, e.g., the intensity of sound as a function of the amplitude of the sound waves), or (c) a grabbag of descriptive studies (see Melzack, 1973, Ch. 2). Of these, only psychophysics could be construed as being about qualia per se. And it is obvious that psychophysics touches only the *functional* aspect of sensation, not its qualitative character. Psychophysical experiments done on you would have the same results if done on any system Psychofunctionally equivalent to you, even if it had inverted or absent qualia. If experimental results would be unchanged whether or not the experimental subjects have inverted or absent qualia, they can hardly be expected to cast light on the nature of qualia.

Indeed, on the basis of the kind of conceptual apparatus now available in psychology, I do not see how psychology in anything like its present incarnation *could* explain qualia. We cannot now conceive how psychology could explain qualia, though we *can* conceive how psychology could explain believing, desiring, hoping, etc. (see Fodor, 1975). That something is currently inconceivable is not a good reason to think it is impossible. Concepts could be developed tomorrow that would make what is now inconceivable conceivable. But all we have to go on is what we know, and on the basis of what we have to go on, it looks as if qualia are not in the domain of psychology.

Objection: If the Psychofunctional simulation just described has the same beliefs I have, then among its beliefs will be the belief that it now has a headache (since I now am aware of having a headache). But then you must say that its belief is mistaken—and how can such a belief be mistaken?

Reply: The objection evidently assumes some version of the Incorrigibility Thesis (if x believes he has a pain, it follows that he does have a pain). I believe the Incorrigibility Thesis to be false. But even if it is true, it is a double-edged sword. For one can just as well use it to argue that Psychofunctionalism's difficulties with qualia infect its account of belief too. For if the homunculi-headed simulation is in a state Psychofunctionally equivalent to believing it is in pain, yet has no qualia, and hence no pain, then if the Incorrigibility Thesis is true, it does not believe it is in pain either. But if it is in a state Psychofunctionally equivalent to belief without believing, belief is not a Psychofunctional state.

Objection: At one time it was inconceivable that temperature could be a property of matter, if matter was composed only of particles bouncing about; but it would not have been rational to conclude temperature was not in the domain of physics. Reply: First, what the objection says was inconceivable was probably never inconceivable. When the scientific community could conceive of matter as bouncing particles, it

could probably also conceive of heat as something to do with the motion of the particles. Bacon's theory that heat was motion was introduced at the inception of theorizing about heat—a century before Galileo's primitive precursor of a thermometer, and even before distinctions among the temperature of x, the perceived temperature of x, and x's rate of heat conduction were at all clear (Kuhn, 1961). Second, there is quite a difference between saying something is not in the domain of physics and saying something is not in the domain of psychology. Suggesting that temperature phenomena are not in the domain of physics is suggesting that they are not explainable at all.

It is no objection to the suggestion that qualia are not psychological entities that qualia are the very paradigm of something in the domain of psychology. As has often been pointed out, it is in part an empirical question what is in the domain of any particular branch of science. The liquidity of water turns out not to be explainable by chemistry, but rather by subatomic physics. Branches of science have at any given time a set of phenomena they seek to explain. But it can be discovered that some phenomenon which seemed central to a branch of science is actually in the purview of a different branch.

Suppose psychologists discover a *correlation* between qualitative states and certain cognitive processes. Would that be any reason to think the qualitative states are identical to the cognitive states they are correlated with? Certainly not. First, what reason would there be to think this correlation would hold in the homunculi-headed systems that Psychofunctionally simulate us? Second, although a case can be made that certain sorts of general correlations between Fs and Gs provide reason to think F is G, this is only the case when the predicates are predicates of different theories, one of which is reducible to the other. For example, there is a correlation between thermal and electrical conductivity (asserted by the Wiedemann-Franz Law), but it would be silly to suggest that this shows thermal conductivity is electrical conductivity (see Block, 1971, Ch. 3).

I know of only one serious attempt to fit ''consciousness'' into information-flow psychology: the program in Dennett, 1978. But Dennett fits consciousness into information-flow psychology only by claiming that the contents of consciousness are exhausted by judgments. His view is that to the extent that qualia are not judgments (or beliefs), they are spurious theoretical entities that we postulate to explain why we find ourselves wanting to say all sorts of things about what is going on in our minds.

Dennett's doctrine has the relation to qualia that the U.S. Air Force had to so many Vietnamese villages: he destroys qualia in order to save them. Is it not more reasonable to tentatively hypothesize that qualia are determined by the physiological or physicochemical nature of our information processing, rather than by the information flow per se?

The Absent Qualia Argument exploits the possibility that the Functional or Psychofunctional state Functionalists or Psychofunctionalists would want to identify with pain can occur without any quale occurring. It also seems to be conceivable that the latter occur without the former. Indeed, there are facts that lend plausibility to this view. After frontal lobotomies, patients typically report that they still have pains, though the pains no longer bother them (Melzack, 1973, p. 95). These patients show all the ''sensory'' signs of pain (e.g., recognizing pin pricks as sharp), but they often have little or no desire to avoid ''painful'' stimuli.

One view suggested by these observations is that each pain is actually a *composite* state whose components are a quale and a Functional or Psychofunctional state.[14] Or what amounts to much the same idea, each pain is a quale playing a certain Functional or Psychofunctional role. If this view is right, it helps to explain how people can have believed such different theories of the nature of pain and other sensations: they have emphasized one component at the expense of the other. Proponents of behaviorism and functionalism have had one component in mind; proponents of private ostensive definition have had the other in mind. Both approaches err in trying to give one account of something that has two components of quite different natures.

3.0 Chauvinism vs. Liberalism

It is natural to understand the psychological theories Psychofunctionalism adverts to as theories of *human* psychology. On Psychofunctionalism, so understood, it is logically impossible for a system to have beliefs, desires, etc., except insofar as psychological theories true of us are true of it. Psychofunctionalism (so understood) stipulates that Psychofunctional equivalence to us is necessary for mentality.

But even if Psychofunctional equivalence to us is a condition on our *recognition of mentality*, what reason is there to think it is a condition on mentality itself? Could there not be a wide variety of possible psychological processes that can underlie mentality, of which we instantiate only one type? Suppose we meet Martians and find that they are roughly Functionally (but not Psychofunctionally) equivalent to us. When we get to know Martians, we find them about as different from us as humans we know. We develop extensive cultural and commercial intercourse with them. We study each other's science and philosophy journals, go to each other's movies, read each other's novels, etc. Then Martian and Earthian psychologists compare notes, only to find that in underlying psychology, Martians and Earthians are very different. They soon agree that the difference can be described as follows. Think of humans and Martians as if they were products of conscious design. In any such design project, there will be various options. Some capacities can be built in (innate), others learned. The brain can be designed to accomplish tasks using as much memory capacity as necessary in order to

minimize use of computation capacity; or, on the other hand, the designer could choose to conserve memory space and rely mainly on computation capacity. Inferences can be accomplished by systems which use a few axioms and many rules of inference, or, on the other hand, few rules and many axioms. Now imagine that what Martian and Earthian psychologists find when they compare notes is that Martians and Earthians differ as if they were the end products of maximally different design choices (compatible with rough Functional equivalence in adults). Should we reject our assumption that Martians can enjoy our films, believe their own apparent scientific results, etc.? Should they ''reject'' their ''assumption'' that we ''enjoy'' their novels, ''learn'' from their textbooks, etc.? Perhaps I have not provided enough information to answer this question. After all, there may be many ways of filling in the description of the Martian-human differences in which it would be reasonable to suppose there simply is no fact of the matter, or even to suppose that the Martians do not deserve mental ascriptions. But surely there are many ways of filling in the description of the Martian-Earthian difference I sketched on which it would be perfectly clear that even if Martians behave differently from us on subtle psychological experiments, they nonetheless think, desire, enjoy, etc. To suppose otherwise would be crude human chauvinism. (Remember theories are chauvinist insofar as they falsely *deny* that systems have mental properties and liberal insofar as they falsely *attribute* mental properties.)

So it seems as if in preferring Psychofunctionalism to Functionalism, we erred in the direction of human chauvinism. For if mental states are Psychofunctional states, and if Martians do not have these Psychofunctional states, then they do not have mental states either. In arguing that the original homunculi-headed simulations (taken as Functional simulations) had no mentality, I appealed, in effect, to the following principle: if the sole reason to think system x has mentality is that x was built to be Functionally equivalent to us, then differences between x and us in underlying information processing and/or neurophysiology are reasons to doubt whether x has mental states. But this principle does not dictate that a system can have mentality only insofar as it is Psychofunctionally equivalent to us. Psychofunctional equivalence to us is a sufficient condition for at least those aspects of mentality in the domain of psychology, but it is not obvious that it is a necessary condition of any aspects of mentality.

An obvious suggestion of a way out of this difficulty is to identify mental states with Psychofunctional states, taking the domain of psychology to include *all creatures with mentality*, including Martians. The suggestion is that we define ''Psychofunctionalism'' in terms of ''universal'' or ''cross-system'' psychology, rather than the human psychology I assumed earlier. Universal psychology, however, is a suspect discipline. For how are we to decide what systems should be included in the *domain* of universal psychology? One possible way of deciding what systems have mentality, and are thus in the

domain of universal psychology, would be to use some *other* developed theory of mentality, e.g., behaviorism or Functionalism. But such a procedure would be at least as ill-justified as the other theory used. Further, if Psychofunctionalism must presuppose some other theory of mind, we might just as well accept the other theory of mind instead.

Perhaps universal psychology will avoid this ''domain'' problem in the same way other branches of science avoid it or seek to avoid it. Other branches of science start with tentative domains based on intuitive and prescientific versions of the concepts the sciences are supposed to explicate. They then attempt to develop natural kinds in a way which allows the formulations of lawlike generalizations which apply to all or most of the entities in the prescientific domains. In the case of many branches of science—including biological and social sciences such as genetics and linguistics—the prescientific domain turned out to be suitable for the articulation of lawlike generalizations.

Now it may be that we shall be able to develop universal psychology in much the same way we develop Earthian psychology. We decide on an intuitive and prescientific basis what creatures to include in its domain, and work to develop natural kinds of psychological theory which apply to all or at least most of them. Perhaps the study of a wide range of organisms found on different worlds will one day lead to theories that determine truth conditions for the attribution of mental states like belief, desire, etc., applicable to systems which are pretheoretically quite different from us. Indeed, such cross-world psychology will no doubt require a whole new range of mentalistic concepts. Perhaps there will be families of concepts corresponding to belief, desire, etc., that is, a family of belieflike concepts, desirelike concepts, etc. If so, the universal psychology we develop shall, no doubt, be somewhat dependent on which new organisms we discover first. Even if universal psychology is in fact possible, however, there will certainly be many possible organisms whose mental status is indeterminate.

On the other hand, it may be that universal psychology is *not* possible. Perhaps life in the universe is such that we shall simply have no basis for reasonable decisions about what systems are in the domain of psychology and what systems are not.

If universal psychology *is* possible, the problem I have been raising vanishes. Universal-Psychofunctionalism avoids the liberalism of Functionalism and the chauvinism of human-Psychofunctionalism. But the question of whether universal psychology is possible is surely one which we have no way of answering now.

Here is a summary of the argument so far:

1. Functionalism has the bizarre consequence that a homunculi-headed simulation of you has qualia. This puts the burden of proof on the Functionalist to give us some reason for believing his doctrine. However, the one argument for Functionalism in the literature is no good, and so Functionalism shows no sign of meeting the burden of proof.

2. Psychofunctional simulations of us share whatever states are in the domain of psychology, so the Psychofunctional homunculi-head does not cast doubt on Psychofunctional theories of cognitive states, but only on Psychofunctionalist theories of qualia, there being a doubt as to whether qualia are in the domain of psychology.
3. Psychofunctionalist theories of mental states that are in the domain of psychology, however, are hopelessly chauvinist.

So one version of Functionalism has problems with liberalism; the other has problems with chauvinism. As to qualia, if they are in the domain of psychology, then Psychofunctionalism with respect to qualia is just as chauvinist as Psychofunctionalism with respect to belief. On the other hand, if qualia are not in the domain of psychology, the Psychofunctionalist homunculi-head can be used against Psychofunctionalism with respect to qualia. For the only thing that shields Psychofunctionalism with respect to mental state S from the homunculi-head argument is that if you have S, then any Psychofunctional simulation of you must have S, because the correct theory of S applies to it just as well as to you.

3.1 The Problem of the Inputs and the Outputs

I have been supposing all along (as Psychofunctionalists often do—see Putnam, 1967) that inputs and outputs can be specified by neural impulse descriptions. But this is a chauvinist claim, since it precludes organisms without neurons (e.g., machines) from having functional descriptions. How can one avoid chauvinism with respect to specification of inputs and outputs? One way would be to characterize the inputs and outputs *only as* inputs and outputs. So the functional description of a person might list outputs by number: output_1, $\text{output}_2, \ldots$ Then a system could be functionally equivalent to you if it had a set of states, inputs, and outputs causally related to one another in the way yours are, no matter what the states, inputs, and outputs were like. Indeed, though this approach violates the demand of some functionalists that inputs and outputs be physically specified, other functionalists—those who insist only that input and output descriptions be *nonmental*—may have had something like this in mind. This version of functionalism does not "tack down" functional descriptions at the periphery with relatively specific descriptions of inputs and outputs; rather, this version of functionalism treats inputs and outputs just as all versions of functionalism treat internal states. That is, this version specifies states, inputs, and outputs only by requiring that they *be* states, inputs, and outputs.

The trouble with this version of functionalism is that it is wildly liberal. Economic systems have inputs and outputs, e.g., influx and outflux of credits and debits. And economic systems also have a rich variety of internal states, e.g., having a rate of increase of GNP equal to double the Prime Rate. It does not seem impossible that a

wealthy sheik could gain control of the economy of a small country, e.g., Bolivia, and manipulate its financial system to make it functionally equivalent to a person, e.g., himself. If this seems implausible, remember that the economic states, inputs, and outputs designated by the sheik to correspond to his mental states, inputs, and outputs need not be ''natural'' economic magnitudes. Our hypothetical sheik could pick *any* economic magnitudes at all—e.g., the fifth time derivative of the balance of payments. His only constraint is that the magnitudes he picks be economic, that their having such and such values be inputs, outputs, and states, and that he be able to set up a financial structure which can be made to fit the intended formal mold. The mapping from psychological magnitudes to economic magnitudes could be as bizarre as the sheik requires.

This version of functionalism is far too liberal and must therefore be rejected. If there are any fixed points when discussing the mind-body problem, one of them is that the economy of Bolivia could not have mental states, no matter how it is distorted by powerful hobbyists. Obviously, we must be more specific in our descriptions of inputs and outputs. The question is: is there a description of inputs and outputs specific enough to avoid liberalism, yet general enough to avoid chauvinism? I doubt that there is.

Every proposal for a description of inputs and outputs I have seen or thought of is guilty of either liberalism or chauvinism. Though this paper has concentrated on liberalism, chauvinism is the more pervasive problem. Consider standard Functional and Psychofunctional descriptions. Functionalists tend to specify inputs and outputs in the manner of behaviorists: outputs in terms of movements of arms and legs, sound emitted and the like; inputs in terms of light and sound falling on the eyes and ears. Such descriptions are blatantly *species-specific*. Humans have arms and legs, but snakes do not—and whether or not snakes have mentality, one can easily imagine snake-like creatures that do. Indeed, one can imagine creatures with all manner of input-output devices, e.g., creatures that communicate and manipulate by emitting strong magnetic fields. Of course, one could formulate Functional descriptions for each such species, and somewhere in disjunctive heaven there is a disjunctive description which will handle all species that ever actually exist in the universe (the description may be infinitely long). But even an appeal to such suspicious entities as infinite disjunctions will not bail out Functionalism, since even the amended view will not tell us what there is in common to pain-feeling organisms in virtue of which they all have pain. And it will not allow the ascription of pain to some hypothetical (but nonexistent) pain-feeling creatures. Further, these are just the grounds on which functionalists typically acerbically reject the disjunctive theories sometimes advanced by desperate physicalists. If functionalists suddenly smile on wildly disjunctive states to save themselves from chauvinism, they will have no way of defending themselves from physicalism.

Standard Psychofunctional descriptions of inputs and outputs are also species-specific (e.g., in terms of neural activity) and hence chauvinist as well.

The chauvinism of standard input-output descriptions is not hard to explain. The variety of possible intelligent life is enormous. Given any fairly specific descriptions of inputs and outputs, any high-school-age science-fiction buff will be able to describe a sapient sentient being whose inputs and outputs fail to satisfy that description.

I shall argue that *any physical description* of inputs and outputs (recall that many functionalists have insisted on physical descriptions) yields a version of functionalism that is inevitably chauvinist or liberal. Imagine yourself so badly burned in a fire that your optimal way of communicating with the outside world is via modulations of your EEG pattern in Morse Code. You find that thinking an exciting thought produces a pattern that your audience agrees to interpret as a dot, and a dull thought produces a ''dash.'' Indeed, this fantasy is not so far from reality. According to a recent newspaper article (*Boston Globe*, March 21, 1976), ''at UCLA scientists are working on the use of EEG to control machines.... A subject puts electrodes on his scalp, and thinks an object through a maze.'' The ''reverse'' process is also presumably possible: others communicating with you in Morse Code by producing bursts of electrical activity that affect your brain (e.g., causing a long or short afterimage). Alternatively, if the cerebroscopes that philosophers often fancy become a reality, your thoughts will be readable directly from your brain. Again, the reverse process also seems possible. In these cases, *the brain itself becomes an essential part of one's input and output devices*. This possibility has embarrassing consequences for functionalists. You will recall that functionalists pointed out that physicalism is false because a single mental state can be realized by an indefinitely large variety of physical states that have no necessary and sufficient physical characterization. But if this functionalist point against physicalism is right, *the same point applies to inputs and outputs*, since the physical realization of mental states can serve as an essential part of the input and output devices. That is, on any sense of 'physical' in which the functionalist criticism of physicalism is correct, *there will be no physical characterization that applies to all and only mental systems' inputs and outputs*. Hence, any attempt to formulate a functional description with physical characterizations of inputs and outputs will inevitably either exclude some systems with mentality or include some systems without mentality. Hence, *functionalists cannot avoid both chauvinism and liberalism*.

So physical specifications of inputs and outputs will not do. Moreover, mental or ''action'' terminology (e.g., ''punching the offending person'') can not be used either, since to use such specifications of inputs or outputs would be to give up the functionalist program of characterizing mentality in nonmental terms. On the other hand, as you will recall, characterizing inputs and outputs simply *as* inputs and outputs is inevitably liberal. I, for one, do not see how there can be a vocabulary for describing inputs and outputs that avoids both liberalism and chauvinism. I do not claim that this is a conclusive argument against functionalism. Rather, like the functionalist argument against physicalism, it is best construed as a burden-of-proof argument. The function-

alist says to the physicalist: ''It is very hard to see how there could be a single physical characterization of the internal states of all and only creatures with mentality.'' I say to the functionalist: ''It is very hard to see how there could be a single physical characterization of the inputs and outputs of all and only creatures with mentality.'' In both cases, enough has been said to make it the responsibility of those who think there could be such characterizations to sketch how they could be possible.[15]

Notes

Reprinted with revision and abridgment from C. W. Savage, ed., *Minnesota Studies in the Philosophy of Science*, vol. 9, 261–325 (Minneapolis: University of Minnesota Press, 1978). Reprinted in N. Block, ed., *Readings in Philosophy of Psychology*, vol. 1 (Cambridge, MA: Harvard University Press, 1980). Reprinted (shortened version) in W. Lycan, ed., *Mind and Cognition*, 444–469 (Oxford: Blackwell, 1990). Reprinted (shortened version) in D. M. Rosenthal, ed., *The Nature of Mind*, 211–229 (Oxford: Oxford University Press, 1991). Reprinted (shortened version) in B. Beakley and P. Ludlow, eds., *Philosophy of Mind* (Cambridge, MA: MIT Press, 1992). Reprinted (shortened version) in Alvin Goldman, ed., *Readings in Philosophy and Cognitive Science* (Cambridge, MA: MIT Press, 1993). German translation, ''Schwierigkeiten mit dem Funktionalismus,'' in D. Münch, Hg., *Kognitionswissenschaft: Grundlagen, Probleme, Perspektiven*, 159–225 (Frankfurt: Suhrkamp, 1992). French translation, ''Le fonctionnalisme face au problème des qualia,'' in *Les Etudes Philosophiques* 3: 337–369, 1992. Spanish translation in E. Rabossi, ed., *Filosofia y Cienca Cognitiva* (Buenos Aires and Barcelona: Libreria Paidós, 1996). Reprinted (much shortened version) in W. Lycan, ed., *Mind and Cognition*, 2nd ed. (Oxford: Blackwell, 1999). Reprinted in Jack Crumley, ed., *Problems in Mind: Readings in Contemporary Philosophy of Mind* (Mayfield, 1999). Reprinted in David Chalmers, ed., *Philosophy of Mind: Classical and Contemporary Readings* (Oxford: Oxford University Press, 2002). Reprinted in *Philosophy of Mind: Contemporary Readings*, a volume in the Routledge ''Contemporary Readings'' series edited by Timothy William O'Connor and David Robb (New York: Routledge, 2003).

1. See Fodor, 1965, 1968a; Lewis, 1966, 1972; Putnam, 1966, 1967, 1970, 1975a; Armstrong, 1968; Locke, 1968; perhaps Sellars, 1968; perhaps Dennett, 1969, 1978b; Nelson, 1969, 1975 (but see also Nelson, 1976); Pitcher, 1971; Smart, 1971; Block & Fodor, 1972; Harman, 1973; Lycan, 1974; Grice, 1975; Shoemaker, 1975; Wiggins, 1975; Field, 1978.

2. The converse is also true.

3. Indeed, if one defines 'behaviorism' as the view that mental terms can be defined in nonmental terms, then functionalism *is* a version of behaviorism.

4. State type, not state token. Throughout the chapter, I shall mean by 'physicalism' the doctrine that says each distinct type of mental state is identical to a distinct type of physical state; for example, pain (the universal) is a physical state. Token physicalism, on the other hand, is the (weaker) doctrine that each particular datable pain is a state of some physical type or other. Functionalism shows that type physicalism is false, but it does not show that token physicalism is false.

By 'physicalism,' I mean *first-order* physicalism, the doctrine that, e.g., the property of being in pain is a first-order (in the Russell-Whitehead sense) physical property. (A first-order property is one whose definition does not require quantification over properties; a second-order property is one whose definition requires quantification over first-order properties—and not other properties.) The claim that being in pain is a second-order physical property is actually a (physicalist) form of functionalism. See Putnam, 1970.

'Physical property' could be defined for the purposes of this chapter as a property expressed by a predicate of some true physical theory or, more broadly, by a predicate of some true theory of physiology, biology, chemistry, or physics. Of course, such a definition is unsatisfactory without characterizations of these branches of science (see Hempel, 1970, for further discussion). This problem could be avoided by characterizing 'physical property' as: property expressed by a predicate of some true theory adequate for the explanation of the phenomena of nonliving matter. I believe that the difficulties of this account are about as great as those of the previous account. Briefly, it is conceivable that there are physical laws that ''come into play'' in brains of a certain size and complexity, but that nonetheless these laws are ''translatable'' into physical language, and that, so translated, they are clearly physical laws (though irreducible to other physical laws). Arguably, in this situation, physicalism could be true—though not according to the account just mentioned of 'physical property'.

Functionalists who are also physicalists have formulated broadly physicalistic versions of functionalism. As functionalists often point out (Putnam, 1967), it is logically possible for a given abstract functional description to be satisfied by a nonphysical object, e.g., a soul. One can formulate a physicalistic version of functionalism simply by explicitly ruling out this possibility. One such physicalistic version of functionalism is suggested by Putnam (1970), Field (1978), and Lewis (in conversation): having pain is identified with a second-order physical property, a property that consists of having certain first-order physical properties if certain other first-order physical properties obtain. This doctrine combines functionalism (which can be formulated as the doctrine that having pain is the property of having certain properties if certain other properties obtain) with token physicalism. Of course, the Putnam-Lewis-Field doctrine is *not* a version of (first-order) type physicalism; indeed, the P-L-F doctrine is incompatible with (first-order) type physicalism.

5. I mentioned two respects in which Functionalism and Psychofunctionalism differ. First, Functionalism identifies pain with its Ramsey functional correlate with respect to a common-sense psychological theory, and Psychofunctionalism identifies pain with its Ramsey functional correlate with respect to a scientific psychological theory. Second, Functionalism requires common-sense specification of inputs and outputs, and Psychofunctionalism has the option of using empirical-theory construction in specifying inputs and outputs so as to draw the line between the inside and outside of the organism in a theoretically principled way.

I shall say a bit more about the Psychofunctionalism/Functionalism distinction. According to the preceding characterization, Psychofunctionalism and Functionalism are theory relative. That is, we are told not what pain *is*, but, rather, what pain is *with respect to this or that theory*. But Psychofunctionalism can be defined as the doctrine that mental states are constituted by causal relations among whatever psychological events, states, processes, and other entities—as well as inputs and outputs—actually obtain in us in whatever ways those entities are actually causally related

to one another. Therefore, if current theories of psychological processes are correct in adverting to storage mechanisms, list searchers, item comparators, and so forth, Psychofunctionalism will identify mental states with causal structures that involve storage, comparing, and searching processes as well as inputs, outputs, and other mental states.

Psychofunctional equivalence can be similarly characterized without overt relativizing to theory. Let us distinguish between weak and strong equivalence (Fodor, 1968a). Assume we have agreed on some descriptions of inputs and outputs. I shall say that organisms x and y are weakly or behaviorally equivalent if and only if they have the same output for any input or sequence of inputs. If x and y are weakly equivalent, each is a weak simulation of the other. I shall say x and y are *strongly* equivalent relative to some branch of science if and only if (1) x and y are weakly equivalent, and (2) that branch of science has in its domain processes that mediate inputs and outputs, and x's and y's inputs and outputs are mediated by the same combination of weakly equivalent processes. If x and y are strongly equivalent, they are strong simulations of each other.

We can now give a characterization of a Psychofunctional equivalence relation that is not overtly theory relative. This Psychofunctional equivalence relation is strong equivalence with respect to psychology. (Note that 'psychology' here denotes a branch of science, not a particular theory in that branch.)

This Psychofunctional equivalence relation differs in a number of respects from those described earlier. For example, for the sort of equivalence relation described earlier, equivalent systems need not have any common output if they share a given sequence of inputs. In machine terms, the equivalence relations described earlier require only that equivalent systems have a common machine table (of a certain type); the current equivalence relation requires, in addition, that equivalent systems be in the same state of the machine table. This difference can be eliminated by more complex formulations.

Ignoring differences between Functionalism and Psychofunctionalism in their characterizations of inputs and outputs, we can give a very crude account of the Functionalism/Psychofunctionalism distinction as follows: Functionalism identifies mental states with causal structures involving conscious mental states, inputs, and outputs; Psychofunctionalism identifies mental states with the same causal structures, elaborated to include causal relations to *unconscious* mental entities as well. That is, the causal relations adverted to by Functionalism are a subset of those adverted to by Psychofunctionalism. Thus, weak or behavioral equivalence, Functional equivalence, and Psychofunctional equivalence form a hierarchy. All Psychofunctionally equivalent systems are Functionally equivalent, and all Functionally equivalent systems are weakly or behaviorally equivalent.

Although the characteristics of Psychofunctionalism and Psychofunctional equivalence just given are not overtly theory relative, they have the same vagueness problems as the characterizations given earlier. I pointed out that the Ramsey functional-correlate characterizations suffer from vagueness about level of abstractness of psychological theory—e.g., are the psychological theories to cover only humans who are capable of *weltschmerz*, all humans, all mammals, or what? The characterization of Psychofunctionalism just given allows a similar question: what is to count as a psychological entity or process? If the answer is an entity in the domain of some true psychological theory, we have introduced relativity to theory. Similar points apply to the identification of Psychofunctional equivalence, with strong equivalence with respect to psychology.

Appeal to unknown, true psychological theories introduces another kind of vagueness problem. We can allocate current theories among branches of science by appealing to concepts or vocabulary currently distinctive to those branches. But we cannot timelessly distinguish among branches of science by appealing to their distinctive concepts or vocabulary, because we have no idea what concepts and vocabulary the future will bring. If we did know, we would more or less have future theories now. Worse still, branches of science have a habit of coalescing and splitting, so we cannot know whether the science of the future will countenance anything at all like psychology as a branch of science.

One consequence of this vagueness is that no definite answer can be given to the question, Does Psychofunctionalism as I have described it characterize mental states partly in terms of their relations to *neurological* entities? I think the best anyone can say is: at the moment, it seems not. Psychology and neurophysiology seem to be separate branches of science. Of course, it is clear that one must appeal to neurophysiology to explain some psychological phenomena, e.g., how being hit on the head causes loss of language ability. However, it seems as if this should be thought of as ''descending'' to a lower level in the way evolutionary biology appeals to physics (e.g., cosmic rays hitting genes) to partially explain mutation.

6. The basic idea for this example is due to Putnam (1967). I am indebted to many conversations with Hartry Field on the topic. Putnam's attempt to defend functionalism from the problem posed by such examples is discussed in Section 1.4 of this essay.

One potential difficulty for Functionalism is provided by the possibility that one person may have two radically different Functional descriptions of the sort that justify attribution of mentality. In such a case, Functionalists might have to ascribe two radically different systems of belief, desire, etc., to the same person, or suppose that there is no fact of the matter about what the person's propositional attitudes are. Undoubtedly, Functionalists differ greatly on what they make of this possibility, and the differences reflect positions on such issues as indeterminacy of translation.

7. Shoemaker, 1975, argues (in reply to Block and Fodor, 1972) that absent qualia are logically impossible, that is, that it is logically impossible that two systems be in the same functional state yet one's state have and the other's state lack qualitative content. If Shoemaker is right, it is wrong to doubt whether the homunculi-headed system has qualia. I attempt to show Shoemaker's argument to be fallacious in Block, 1980.

8. Since there is a difference between the role of the little people in producing your functional organization in the situation just described and the role of the homunculi in the homunculi-headed simulations this chapter began with, presumably Putnam's condition could be reformulated to rule out the latter without ruling out the former. But this would be a most ad hoc maneuver. Further, there are other counterexamples which suggest that a successful reformulation is likely to remain elusive.

Careful observation of persons who have had the nerve bundle connecting the two halves of the brain (the *corpus callosum*) severed to prevent the spread of epilepsy, suggest that each half of the brain has the functional organization of a sentient being. The same is suggested by the observation that persons who have had one hemisphere removed or anesthetized remain sentient beings. It was once thought that the right hemisphere had no linguistic capacity, but it is now known that the adult right hemisphere has the vocabulary of a 14-year-old and the syntax of a 5-

year-old (*Psychology Today*, 12/75, p. 121). Now the functional organization of each hemisphere is different from the other and from that of a whole human. For one thing, in addition to inputs from the sense organs and outputs to motor neurons, each hemisphere has many input and output connections to the other hemisphere. Nonetheless, each hemisphere may have the functional organization of a sentient being. Perhaps Martians have many more input and output organs than we do. Then each half brain could be functionally like a whole Martian brain. If each of our hemispheres has the functional organization of a sentient being, then a Putnamian proposal would rule us out (except for those of us who have had hemispherectomies) as pain-feeling organisms.

Further, it could turn out that other parts of the body have a functional organization similar to that of some sentient being. For example, perhaps individual neurons have the same functional organization as some species of insect.

(The argument of the last two paragraphs depends on a version of functionalism that construes inputs and outputs as neural impulses. Otherwise, individual neurons could not have the same functional organization as insects. It would be harder to think of such examples if, for instance, inputs were taken to be irradiation of sense organs or the presence of perceivable objects in the ''range'' of the sense organs.)

9. A further indication that our intuitions are in part governed by the neurophysiological and psychological differences between us and the original homunculi-headed simulation (construed as a Functional simulation) is that intuition seems to founder on an intermediate case: a device that simulates you by having a billion little men each of whom simulates one of your neurons. It would be like you in psychological mechanisms, but not in neurological mechanisms, except at a very abstract level of description.

There are a number of differences between the original homunculi-heads and the elementary-particle-people example. The little elementary-particle people were not described as knowing your functional organization or trying to simulate it, but in the original example, the little men have *as their aim* simulating your functional organization. Perhaps when we know a certain functional organization is intentionally produced, we are thereby inclined to regard the thing's being functionally equivalent to a human as a misleading fact. One could test this by changing the elementary-particle-people example so that the little people have the aim of simulating your functional organization by simulating elementary particles; this change seems to me to make little intuitive difference.

There are obvious differences between the two types of examples. It is *you* in the elementary case and the change is *gradual*; these elements seem obviously misleading. But they can be eliminated without changing the force of the example much. Imagine, for example, that your spouse's parents went on the expedition and that your spouse has been made of the elementary-particle people since birth.

10. Compare the first sentence with 'The fish eaten in Boston stank.' The reason it is hard to process is that ''raced'' is naturally read as active rather than passive. See Fodor, Bever, and Garrett, 1974, p. 360. For a discussion of why the second sentence is grammatical, see Fodor and Garrett, 1967; Bever, 1970; and Fodor, Bever, and Garrett, 1974.

11. We often fail to be able to conceive of how something is possible because we lack the relevant theoretical concepts. For example, before the discovery of the mechanism of genetic duplication,

Haldane argued persuasively that no conceivable physical mechanism could do the job. He was right. But instead of urging that scientists should develop ideas that would allow us to conceive of such a physical mechanism, he concluded that a *non*physical mechanism was involved. (I owe the example to Richard Boyd.)

12. One could avoid this difficulty by allowing *names* or demonstratives in one's physical theory. For example, one could identify protons as the particles with such and such properties contained in the nuclei of all atoms of the Empire State Building. No such move will save this argument for Psychofunctionalism, however. First, it is contrary to the idea of functionalism, since functionalism purports to identify mental states with abstract causal structures; one of the advantages of functionalism is that it avoids appeal to ostension in definition of mental states. Second, tying Psychofunctionalism to particular named entities will inevitably result in chauvinism. See Section 3.1.

13. Suppose a man who has good color vision mistakenly uses ''red'' to denote green and ''green'' to denote red. That is, he simply confuses the two words. Since his confusion is purely linguistic, though he says of a green thing that it is red, he does not *believe* that it is red, any more than a foreigner who has confused ''ashcan'' with ''sandwich'' believes people eat ashcans for lunch. Let us say that the person who has confused 'red' and 'green' in this way is a victim of Word Switching.

Now consider a different ailment: having red/green inverting lenses placed in your eyes without your knowledge. Let us say a victim of this ailment is a victim of Stimulus Switching. Like the victim of Word Switching, the victim of Stimulus Switching applies ''red'' to green things and vice versa. But the victim of Stimulus Switching *does* have false color beliefs. If you show him a green patch he says *and believes* that it is red.

Now suppose that a victim of Stimulus Switching suddenly becomes a victim of Word Switching as well. (Suppose as well that he is a lifelong resident of a remote Arctic village, and has no standing beliefs to the effect that grass is green, firehydrants are red, and so forth.) He speaks normally, applying ''green'' to green patches and ''red'' to red patches. Indeed, he is functionally normal. But his *beliefs* are just as abnormal as they were before he became a victim of Word Switching. Before he confused the words ''red'' and ''green,'' he applied ''red'' to a green patch, and mistakenly believed the patch to be red. Now he (correctly) says ''green,'' but his belief is still wrong.

So two people can be functionally the same, yet have incompatible beliefs. Hence, the inverted qualia problem infects belief as well as qualia (though presumably only qualitative belief). This fact should be of concern not only to those who hold functional state identity theories of belief, but also to those who are attracted by Harman-style accounts of meaning as functional role. Our double victim—of Word and Stimulus Switching—is a counterexample to such accounts. For his word 'green' plays the normal role in his reasoning and inference, yet since in saying of something that it ''is green,'' he expresses his belief that it is *red*, he uses 'green' with an abnormal meaning.

14. The quale might be identified with a physico-chemical state. This view would comport with a suggestion Hilary Putnam made in the late '60s in his philosophy of mind seminar. See also Ch. 5 of Gunderson, 1971.

15. I am indebted to Sylvain Bromberger, Hartry Field, Jerry Fodor, David Hills, Paul Horwich, Bill Lycan, Georges Rey, and David Rosenthal for their detailed comments on one or another earlier draft of this paper. Beginning in the fall of 1975, parts of earlier versions were read at Tufts University, Princeton University, the University of North Carolina at Greensboro, and the State University of New York at Binghamton.

References

Armstrong, D. *A materialist theory of mind*. London: Routledge & Kegan Paul, 1968.

Bever, T. The cognitive basis for linguistic structures. In J. R. Hayes (Ed.), *Cognition and the development of language*. New York: Wiley, 1970.

Block, N. Are absent qualia impossible? *Philosophical Review*, 1980, 89(2).

Block, N., and Fodor, J. What psychological states are not. *Philosophical Review*, 1972, 81, 159–81.

Chisholm, Roderick. *Perceiving*. Ithaca: Cornell University Press, 1957.

Cummins, R. Functional analysis. *Journal of Philosophy*, 1975, 72, 741–64.

Davidson, D. Mental events. In L. Swanson and J. W. Foster (Eds.), *Experience and theory*. Amherst, University of Massachusetts Press, 1970.

Dennett, D. *Content and consciousness*. London: Routledge & Kegan Paul, 1969.

Dennett, D. Why the law of effect won't go away. *Journal for the Theory of Social Behavior*, 1975, 5, 169–87.

Dennett, D. Why a computer can't feel pain. In *Synthese* 1978a, 38, 3.

Dennett, D. *Brainstorms*. Montgomery, Vt.: Bradford, 1978b.

Feldman, F. Kripke's argument against materialism. *Philosophical Studies*, 1973, 416–19.

Fodor, J. Explanations in psychology. In M. Black (Ed.), *Philosophy in America*. London: Routledge & Kegan Paul, 1965.

Fodor, J. The appeal to tacit knowledge in psychological explanation. *Journal of Philosophy*, 1968b, 65, 627–40.

Fodor, J. Special sciences. *Synthese*, 1974, 28, 97–115.

Fodor, J. *The language of thought*. New York: Crowell, 1975.

Fodor, J., Bever, T., and Garrett, M. *The psychology of language*. New York: McGraw-Hill, 1974.

Fodor, J., and Garrett, M. Some syntactic determinants of sentential complexity. *Perception and Psychophysics*, 1967, 2, 289–96.

Geach, P. *Mental acts*. London: Routledge & Kegan Paul, 1957.

Gendron, B. On the relation of neurological and psychological theories: A critique of the hardware thesis. In R. C. Buck and R. S. Cohen (Eds.), *Boston studies in the philosophy of science VIII*. Dordrecht: Reidel, 1971.

Grice, H. P. Method in philosophical psychology (from the banal to the bizarre). *Proceedings and Addresses of the American Philosophical Association*, 1975.

Gunderson, K. *Mentality and machines*. Garden City: Doubleday Anchor, 1971.

Harman, G. *Thought*. Princeton: Princeton University Press, 1973.

Hempel, C. Reduction: Ontological and linguistic facets. In S. Morgenbesser, P. Suppes & M. White (Eds.), *Essays in honor of Ernest Nagel*. New York: St. Martin's Press, 1970.

Kalke, W. What is wrong with Fodor and Putnam's functionalism? *Nous*, 1969, 3, 83–93.

Kim, J. Phenomenal properties, psychophysical laws, and the identity theory. *The Monist*, 1972, 56(2), 177–92.

Lewis, D. Psychophysical and theoretical identifications. *Australasian Journal of Philosophy*, 1972, 50(3), 249–58.

Locke, D. *Myself and others*. Oxford: Oxford University Press, 1968.

Melzack, R. *The puzzle of pain*. New York: Basic Books, 1973.

Minsky, M. *Computation*. Englewood Cliffs: Prentice-Hall, 1967.

Mucciolo, L. F. The identity thesis and neuropsychology. *Nous*, 1974, 8, 327–42.

Nagel, T. The boundaries of inner space. *Journal of Philosophy*, 1969, 66, 452–58.

Nagel, T. Armstrong on the mind. *Philosophical Review*, 1970, 79, 394–403.

Nagel, T. Review of Dennett's *Content and consciousness*. *Journal of Philosophy*, 1972, 50, 220–34.

Nagel, T. What is it like to be a bat? *Philosophical Review*, 1974, 83, 435–50.

Nelson, R. J. Behaviorism is false. *Journal of Philosophy*, 1969, 66, 417–52.

Nelson, R. J. Behaviorism, finite automata & stimulus response theory. *Theory and Decision*, 1975, 6, 249–67.

Nelson, R. J. Mechanism, functionalism, and the identity theory. *Journal of Philosophy*, 1976, 73, 365–86.

Oppenheim, P. and Putnam, H. Unity of science as a working hypothesis. In H. Feigl, M. Scriven and G. Maxwell (Eds.), *Minnesota studies in the philosophy of science II*. Minneapolis: University of Minnesota Press, 1958.

Pitcher, G. *A theory of perception*. Princeton: Princeton University Press, 1971.

Putnam, H. Brains and behavior. 1963. Reprinted as are all Putnam's articles referred to here (except ''On properties'') in *Mind, language and reality: Philosophical papers*, Vol. 2. London: Cambridge University Press, 1975.

Putnam, H. The mental life of some machines. 1966.

Putnam, H. The nature of mental states (originally published under the title *Psychological Predicates*). 1967.

Putnam, H. On properties. In *Mathematics, matter and method: Philosophical papers*, Vol. 1. London: Cambridge University Press, 1970.

Putnam, H. Philosophy and our mental life. 1975a.

Putnam, H. The meaning of 'meaning'. 1975b.

Rorty, R. Functionalism, machines and incorrigibility. *Journal of Philosophy*, 1972, 69, 203–20.

Scriven, M. *Primary philosophy*. New York: McGraw-Hill, 1966.

Sellars, W. Empiricism and the philosophy of mind. In H. Feigl & M. Scriven (Eds.), *Minnesota studies in philosophy of science I*. Minneapolis: University of Minnesota Press, 1956.

Sellars, W. *Science and metaphysics*. (Ch. 6). London: Routledge & Kegan Paul, 1968.

Shoemaker, S. Functionalism and qualia. *Philosophical studies*, 1975, 27, 271–315.

Shoemaker, S. Embodiment and behavior. In A. Rorty (Ed.), *The identities of persons*. Berkeley: University of California Press, 1976.

Shallice, T. Dual functions of consciousness. *Psychological Review*, 1972, 79, 383–93.

Smart, J. J. C. Reports of immediate experience. *Synthese*, 1971, 22, 346–59.

Wiggins, D. Identity, designation, essentialism, and physicalism. *Philosophia*, 1975, 5, 1–30.