2 What Is Functionalism?

It is doubtful whether doctrines known as ''functionalism'' in fields as disparate as anthropology, literary criticism, psychology, and philosophy of psychology have anything in common but the name. Even in philosophy of psychology, the term is used in a number of distinct senses. The functionalisms of philosophy of psychology are, however, a closely knit group; indeed, they appear to have a common origin in the works of Aristotle (see Hartman, 1977, especially chap. 4).

Three functionalisms have been enormously influential in philosophy of mind and psychology:

Functional Analysis

In this sense of the term, functionalism is a type of explanation and, derivatively, a research strategy, the research strategy of looking for explanations of that type. A functional explanation is one that relies on a decomposition of a system into its component parts; it explains the working of the system in terms of the capacities of the parts and the way the parts are integrated with one another. For example, we can explain how a factory can produce refrigerators by appealing to the capacities of the various assembly lines, their workers and machines, and the organization of these components. The article by Robert Cummins (1975) describes functionalism in this sense. (See also Fodor, 1965, 1968a, 1968b; Dennett, 1975.)

Computation-Representation Functionalism

In this sense of the term, ''functionalism'' applies to an important special case of functional explanation as defined above, namely, to psychological explanation seen as akin to providing a computer program for the mind. Whatever mystery our mental life may initially seem to have is dissolved by functional analysis of mental processes to the point where they are seen to be composed of computations as mechanical as the primitive operations of a digital computer—processes so stupid that appealing to them in psychological explanations involves no hint of question-begging. The key notions of

functionalism in this sense are representation and computation. Psychological states are seen as systematically representing the world via a language of thought, and psychological processes are seen as computations involving these representations. Functionalism in this sense of the term is not explored here but is discussed in volume 2, part one, ''Mental Representation.''

Metaphysical Functionalism

The last functionalism, the one that this part is mainly about, is a theory of *the nature of the mind*, rather than a theory of psychological explanation. Metaphysical functionalists are concerned not with how mental states account for behavior, but rather with what they *are*. The functionalist answer to ''What are mental states?'' is simply that mental states are functional states. Thus theses of metaphysical functionalism are sometimes described as functional state identity theses. The main concern of metaphysical functionalism is the same as that of behaviorism and physicalism. All three doctrines address themselves to such questions as ''What is pain?''—or at least to ''What is there in common to all pains in virtue of which they are pains?''

It is important to note that metaphysical functionalism is concerned (in the first instance) with mental state *types*, not tokens—with *pain*, for instance, and not with particular *pains*. Most functionalists are willing to allow that each *particular* pain is a physical state or event, and indeed that for each type of pain-feeling organism, there is (perhaps) a single type of physical state that realizes pain in that type of organism. Where functionalists differ with physicalists, however, is with respect to the question of what is common to all pains in virtue of which they are pains. The functionalist says the something in common is functional, while the physicalist says it is physical (and the behaviorist says it is behavioral).[1] Thus, in one respect, the disagreement between functionalists and physicalists (and behaviorists) is *metaphysical without being ontological*. Functionalists can be physicalists in allowing that all the entities (things, states, events, and so on) that exist are physical entities, denying only that what binds certain types of things together is a physical property.

Metaphysical functionalists characterize mental states in terms of their causal roles, particularly, in terms of their causal relations to sensory stimulations, behavioral outputs, and other mental states. Thus, for example, a metaphysical functionalist theory of pain might characterize pain in part in terms of its tendency to be caused by tissue damage, by its tendency to cause the desire to be rid of it, and by its tendency to produce action designed to separate the damaged part of the body from what is thought to cause the damage.

What I have said about metaphysical functionalism so far is rather vague, but, as will become clear, disagreements among metaphysical functionalists preclude easy characterization of the doctrine. Before going on to describe metaphysical functionalism in more detail, I shall briefly sketch some of the connections among the functionalist

doctrines just enumerated. One connection is that functionalism in all the senses described has something to do with the notion of a Turing machine (described in the next section). Metaphysical functionalism often identifies mental states with Turing machine ''table states'' (also described in the next section). Computation-representation functionalism sees psychological explanation as something like providing a computer program for the mind. Its aim is to give a functional analysis of mental capacities broken down into their component mechanical processes. If these mechanical processes are *algorithmic*, as is sometimes assumed (without much justification, in my view) then they will be Turing-computable as well (as the Church-Turing thesis assures us).[2] Functional analysis, however, is concerned with the notion of a Turing machine mainly in that providing something like a computer program for the mind is a special case of functional analysis.

Another similarity among the functionalisms mentioned is their relation to physical characterizations. The causal structures with which metaphysical functionalism identifies mental states are realizable by a vast variety of physical systems. Similarly, the information processing mechanisms postulated by a particular computation-representation functionalist theory could be realized hydraulically, electrically, or even mechanically. Finally, functional analysis would normally characterize a manufacturing process abstractly enough to allow a wide variety of types of machines (wood or metal, steam-driven or electrical), workers (human or robot or animal), and physical setups (a given number of assembly lines or half as many dual-purpose assembly lines). A third similarity is that each type of functionalism described legitimates at least one notion of functional equivalence. For example, for functional analysis, one sense of functional equivalence would be: has capacities that contribute in similar ways to the capacities of a whole.

In what follows, I shall try to give the reader a clearer picture of metaphysical functionalism. (''Functionalism'' will be used to mean metaphysical functionalism in what follows.)

Machine Versions of Functionalism

Some versions of functionalism are couched in terms of the notion of a Turing machine, while others are not. A Turing machine is specified by two functions: one from inputs and states to outputs, and one from inputs and states to states. A Turing machine has a finite number of states, inputs, and outputs, and the two functions specify a set of conditionals, one for each combination of state and input. The conditionals are of this form: if the machine is in state S and receives input I, it will then emit output O and go into next state S'. This set of conditionals is often expressed in the form of a machine table (see below). Any system that has a set of inputs, outputs, and states related in the way specified by the machine table is *described* by the machine table

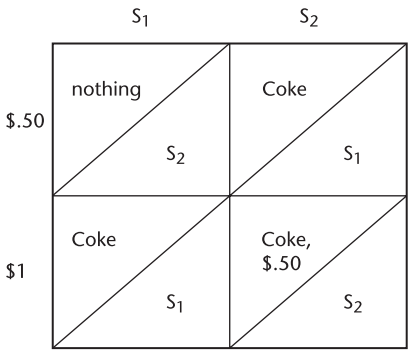

Figure 2.1

and is a *realization* of the abstract automaton specified by the machine table. (This definition actually characterizes a finite transducer, which is just one kind of Turing machine.)

One very simple version of machine functionalism states that each system that has mental states is described by at least one Turing machine table of a certain specifiable sort; it also states that each type of mental state of the system is identical to one of the machine table states specified in the machine table (see Putnam, 1967; Block and Fodor, 1972). Consider, for example, the Turing machine described in the ''Coke machine'' machine table in figure 2.1 (compare Nelson, 1975).

One can get a crude picture of the simple version of machine functionalism described above by considering the claim that S_1 = \$1-desire, and S_2 = \$.50-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 machine table.

Machine versions of functionalism are useful for many purposes, but they do not provide the most general characterization of functionalism. One can achieve more generality by characterizing functionalism as the view that what makes a pain a pain (and, generally, what makes any mental state the mental state it is) is its having a certain causal role.[3] But this formulation buys generality at the price of vagueness. A more precise formulation can be introduced as follows.[4] Let T be a psychological theory (of either common sense or scientific psychology) that tells us (among other things) the relations among pain, other mental states, sensory inputs, and behavioral outputs. Reformulate T so that it is a single conjunctive sentence with all mental state terms as singular terms; for example, ''is angry'' becomes ''has anger.'' Let T so reformulated be written as

$T(s_1 \ldots s_n)$

where $s_1 \ldots s_n$ are terms that designate mental states. Replace each mental state term with a variable and prefix existential quantifiers to form the Ramsey sentence of the theory

$Ex_1 \ldots x_n T(x_1 \ldots x_n)$.

[The ordinary ''E'' is used here instead of the backward ''E'' as the existential quantifier.] Now, if x_i is the variable that replaced 'pain', we can define 'pain' as follows:

y has pain if and only if $Ex_1 \ldots x_n[T(x_1 \ldots x_n)$ & y has $x_i]$.

That is, one has pain just in case he has a state that has certain relations to other states that have certain relations to one another (and to inputs and outputs; I have omitted reference to inputs and outputs for the sake of simplicity). It will be convenient to think of pain as the property expressed by the predicate ''x has pain,'' that is, to think of pain as the property ascribed to someone in saying that he has pain.[5] Then, relative to theory T, pain can be identified with the property expressed by the predicate

$Ex_1 \ldots x_n[T(x_1 \ldots x_n)$ & y has $x_i]$.

For example, take T to be the ridiculously simple theory that pain is caused by pin pricks and causes worry and the emission of loud noises, and worry, in turn, causes brow wrinkling. The Ramsey sentence of T is

$Ex_1Ex_2(x_1$ is caused by pin pricks and causes x_2 and emission of loud noises & x_2 causes brow wrinkling).

Relative to T, pain is the property expressed by the predicate obtained by adding a conjunct as follows:

$Ex_1Ex_2[(x_1$ is caused by pin pricks and causes x_2 and emission of loud noises & x_2 causes brow wrinkling) & y has $x_1]$.

That is, pain is the property that one has when one has a state that is caused by pin pricks, and causes emission of loud noises, and also causes something else, that, in turn, causes brow wrinkling.

We can make this somewhat less cumbersome by letting an expression of the form ''%xFx'' be a singular term meaning the same as an expression of the form ''the property of being an x such that x is F,'' that is, ''being F.'' So %x(x is bigger than a mouse & x is smaller than an elephant) = being bigger than a mouse and smaller than an elephant. Using this notation, we can say

pain = %yEx_1Ex_2[(x_1 is caused by pin pricks and causes x_2 and emission of loud noises & x_2 causes brow wrinkling) & y has x_1].

rather than saying that pain is the property expressed by the predicate

Ex_1Ex_2[(x_1 is caused by pin pricks and causes x_2 and emission of loud noises & x_2 causes brow wrinkling) & y has x_1].

It may be useful to consider a nonmental example. It is sometimes supposed that automotive terms like ''valve-lifter'' or ''carburetor'' are functional terms. Anything that lifts valves in an engine with a certain organizational structure is a valve-lifter. (''Camshaft,'' on the other hand, is a ''structural'' term, at least relative to ''valve-lifter''; a camshaft is *one* kind of device for lifting valves.)

Consider the ''theory'' that says: ''The carburetor mixes gasoline and air and sends the mixture to the ignition chamber, which, in turn ...'' Let us consider ''gasoline'' and ''air'' to be input terms, and let x_1 replace ''carburetor,'' and x_2 replace ''ignition chamber.'' Then the property of being a carburetor would be

$\%yEx_1 \ldots x_n$[(The x_1 mixes gasoline and air and sends the mixture to the x_2, which, in turn ...) & y is an x_1].

That is, being a carburetor = being what mixes gasoline and air and sends the mixture to something else, which, in turn ...

This identification, and the identification of pain with the property one has when one is in a state that is caused by pin pricks and causes loud noises and also causes something else that causes brow wrinkling, would look less silly if the theories of pain (and carburetion) were more complex. But the essential idea of functionalism, as well as its major weakness, can be seen clearly in the example, albeit rather starkly. Pain is identified with an abstract causal property tied to the real world only via its relations, direct and indirect, to inputs and outputs. The weakness is that it seems so clearly conceivable that something could have that causal property, yet *not be* a pain. This point is discussed in detail in ''Troubles with Functionalism'' (Block, 1978; see Shoemaker, 1975, and Lycan, 1979, for critiques of such arguments).

Functionalism and Behaviorism

Many functionalists (such as David Lewis, D. M. Armstrong, and J. J. C. Smart) consider themselves descendants of behaviorists, who attempted to define a mental state in terms of what behaviors would tend to be emitted in the presence of specified stimuli. E.g., the desire for an ice-cream cone might be identified with a set of dispositions, including the disposition to reach out and grasp an ice-cream cone if one is proffered, other things being equal. But, as functionalist critics have emphasized, the phrase ''other things being equal'' is behavioristically illicit, because it can only be filled in with references to *other mental states* (see Putnam, 1963; the point dates back at least to Chisholm, 1957, chap. 11; and Geach, 1957, p. 8). One who desires an ice-cream cone will be disposed to reach for it only if he *knows* it is an ice-cream cone (and not, in general, if he believes it to be a tube of axle-grease), and only if he does not *think*

that taking an ice-cream cone would conflict with *other desires* of more importance to him (such as the desire to lose weight, avoid obligations, or avoid cholesterol). The final nail in the behaviorist coffin was provided by the well-known ''perfect actor'' family of counterexamples. As Putnam argued in convincing detail (1963), it is possible to imagine a community of perfect actors who, by virtue of lawlike regularities, have exactly the behavioral dispositions envisioned by the behaviorists to be associated with absence of pain, even though they do in fact have pain. This shows that no behavioral disposition is a necessary condition of pain, and an exactly analogous example of perfect pain-pretenders shows that no behavioral disposition is a sufficient condition of pain, either.

Functionalism in all its forms differs from behaviorism in two major respects. First, while behaviorists defined mental states in terms of stimuli and responses, they did not think mental states were *themselves* causes of the responses and effects of the stimuli. Behaviorists took mental states to be ''pure dispositions.'' Gilbert Ryle, for example, emphasized that ''to possess a dispositional property is not to be in a particular state, or to undergo a particular change'' (1949, p. 43). Brittleness, according to Ryle, is not a *cause* of breaking, but merely the fact of breaking easily. Similarly, to attribute pain to someone is not to attribute a cause or effect of anything, but simply to say what he would do in certain circumstances. Behaviorists are fictionalists about the mental, hence they cannot allow that mental states have causal powers. Functionalists, by contrast, claim it to be an advantage of their account that it ''allows experiences to be something real, and so to be the effects of their occasions, and the causes of their manifestations (Lewis, 1966, p. 166). Armstrong says that ''[when I think] it is not simply that I would speak or act if some conditions that are unfulfilled were to be fulfilled. Something is currently going on. Rylean behaviorism denies this, and so it is unsatisfactory'' (chapter 13).

The second difference between functionalism and behaviorism is that functionalists emphasize not just the connections between pain and its stimuli and responses, but also its connections to other mental states. Notice, for example, that any full characterization of S_1 in the machine table above would have to refer to S_2 in one way or another, since it is one of the defining characteristics of S_1 that anything in S_1 goes into S_2 when it receives a nickel input. Another example, recall that the Ramsey sentence formulation identifies pain with

$$\%y\mathrm{E}x_1 \ldots x_n[T(x_1 \ldots x_n)\ \&\ y \text{ has } x_i]$$

where the variable x_i replaced 'pain', and the rest of $x_1 \ldots x_n$ replaced the other mental state terms in T. So the functionalist expression that designates pain includes a specification of the relations between pain and all the other mental states related to it, and to inputs and outputs as well. (The role of inputs and outputs would have been better indicated had I written T as

$T(s_1 \ldots s_n, o_1 \ldots o_m, i_1 \ldots i_k)$

explicitly including terms for inputs and outputs.)

Behaviorism is a vague doctrine, and one that is sometimes defined in a way that would make functionalism a version of behaviorism. Even functionalists have offered definitions of 'behaviorism' that would make functionalists behaviorists. For example, if we defined 'behaviorism' as the doctrine that mental states (such as pain) can be characterized in nonmental terms, versions of functionalism along the lines of the Ramsey sentence version sketched above (held by Lewis, Armstrong, Smart, and Sydney Shoemaker) would qualify as versions of behaviorism (since all of the original mental state terms are replaced by variables in the Ramsey sentence). Many other definitions of 'behaviorism' count functionalism as a type of behaviorism. But it would be ludicrously literal-minded to take such definitions very seriously. Clear and general formulations of functionalism were not available until recently, so standard definitions of behaviorism could hardly be expected to draw the boundaries between behaviorism and functionalism with perfect accuracy. Furthermore, given an explicit definition of behaviorism, logical ingenuity can often disguise a functionalist account so as to fit the definition (see Bealer, 1978; Thomas, 1978, for accomplishments of this rather dubious variety). Definitions of behaviorism that count functionalism as behaviorist are misguided precisely *because* they blur the distinctions between functionalism and behaviorism just sketched. A characterization of pain can hardly be counted as behaviorist if it allows that a system could behave (and be disposed to behave) exactly as if it were in pain in all possible circumstances, yet not be in pain.[6]

Is Functionalism Reductionist?

Functionalists sometimes formulate their claim by saying that mental states can only be characterized in terms of other mental states. For instance, a person desires such and such if he would do so and so if he believed doing so and so will get him such and such, and if he believed doing so and so would not conflict with other desires. This much functionalism brings in no reductionism, but functionalists have rarely stopped there. Most regard mental terms as eliminable *all at once*. Armstrong says, for example, ''The logical dependence of purpose on perception and belief, and of perception and belief upon purpose is not circularity in definition. What it shows is that the corresponding concepts must be introduced *together or not at all*'' (1977, p. 88). Shoemaker 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'' (1975). Lewis is more explicit, using a formulation much like the Ramsey sentence formulation given above, which designates mental states by expressions that do not contain any mental terminology (see 1970, 1972 for details).

The same sort of point applies to machine functionalism. Putnam says, "The S_i, to repeat, are specified only *implicitly* by the description" (1967). In the Coke machine automaton described above, the only antecedently understood terms (other than "emit," "go to," and so on) are the input and output terms, "nickel," "dime," and "Coke." The state terms "S_1" and "S_2" in the Coke machine automaton—as in every Turing machine—are given their content entirely in terms of input and output terms (+ logical terms).

Thus functionalism could be said to reduce mentality to input-output structures (note that S_1 and S_2 can have any natures at all, so long as these natures connect them to one another and to the acceptance of nickels and dimes and disbursement of nickels and Cokes as described in the machine table). But functionalism gives us reduction without elimination. Functionalism is not fictionalist about mentality, for each of the functionalist ways of characterizing mental states in terms of inputs and outputs commits itself to the existence of mental states by the use of quantification over mental states, or some equivalent device.[7]

The Varieties of Functionalism

Thus far, I have characterized functionalism without adverting to any of the confusing disagreements among functionalists. I believe that my characterization is correct, but its application to the writings of some functionalists is not immediately apparent. Indeed, the functionalist literature (or, rather, what is generally, and I think correctly, regarded as the functionalist literature) exhibits some bizarre disagreements, the most surprising of which has to do with the relation between functionalism and physicalism. Some philosophers (Armstrong, 1968, 1977; Lewis, 1966, 1970, 1972; Smart, 1971) take functionalism as showing that physicalism is probably *true*, while others (Fodor, 1965; Putnam, 1966; Block and Fodor, chapter 20) take functionalism as showing that physicalism is probably *false*. This is the most noticeable difference among functionalist writings. I shall argue that the Lewis-Armstrong-Smart camp is mistaken in holding that functionalism supports an interesting version of physicalism, and furthermore, that the functionalist insight that they share with the Putnam-Fodor-Harman camp *does* have the consequence that physicalism is probably false. I shall begin with a brief historical sketch.

While functionalism dates back to Aristotle, in its current form it has two main contemporary sources. (A third source, Wittgenstein's, Sellars's and, later, Harman's views on meaning as conceptual role, has also been influential.)

Source I

Putnam (1960) compared the mental states of a person with the machine table states of a Turing machine. He then rejected any identification of mental states with machine

table states, but in a series of articles over the years he moved closer to such an identification, a pattern culminating in ''Psychological Predicates'' (1967). In this article, Putnam came close to advocating a view—which he defended in his philosophy of mind lectures in the late 1960s—that mental states can be identified with machine table states, or rather disjunctions of machine table states. (See Thomas, 1978, for a defence of roughly this view; see Block and Fodor, 1972, and Putnam, 1975, for a critique of such views.)

Fodor (1965, 1968a) developed a similar view (though it was not couched in terms of Turing machines) in the context of a functional-analysis view of psychological explanation (see Cummins, 1975). Putnam's and Fodor's positions were characterized in part by their opposition to physicalism, the view that each *type* of mental state is a physical state.[8] Their argument is at its clearest with regard to the simple version of Turing machine functionalism described above, the view that pain, for instance, is a machine table state. What physical state could be common to all and only realizations of S_1 of the Coke machine automaton described above? The Coke machine could be made of an enormous variety of materials, and it could operate via an enormous variety of mechanisms; it could even be a ''scattered object,'' with parts all over the world, communicating by radio. If someone suggests a putative physical state common to all and only realizations of S_1, it is a simple matter to dream up a nomologically possible machine that satisfies the machine table but does not have the designated physical state. Of course, it is one thing to *say* this and another thing to prove it, but the claim has such overwhelming prima facie plausibility that the burden of proof is on the critic to come up with reason for thinking otherwise. Published critiques (Kalke, 1969; Gendron, 1971; Kim, 1972; Nelson, 1976; Causey, 1977) have in my view failed to meet this challenge.

If we could formulate a machine table for a human, it would be absurd to identify any of the machine table states with a type of *brain* state, since presumably all manner of brainless machines could be described by that table as well. So if pain is a machine table state, it is not a brain state. It should be mentioned, however, that it is possible to *specify* a sense in which a functional state *F* can be said to be physical. For example, *F* might be said to be physical if every system that in fact has *F* is a physical object, or, alternatively, if every realization of *F* (that is, every state that plays the causal role specified by *F*) is a physical state. Of course, the doctrines of ''physicalism'' engendered by such stipulations should not be confused with the version of physicalism that functionalists have argued against (see note 8).

Jaegwon Kim objects that ''the less the physical basis of the nervous system of some organisms resembles ours, the less temptation there will be for ascribing to them sensations or other phenomenal events'' (1972). But his examples depend crucially on considering creatures whose functional organization is much more primitive than ours. He also points out that ''the mere fact that the physical bases of two nervous systems are

different in material composition or physical organization with respect to a certain scheme of classification does not entail that they cannot be in the same physical state with respect to a different scheme.'' Yet the functionalist does not (or, better, should not) claim that functionalism *entails* the falsity of physicalism, but only that the burden of proof is on the physicalist. Kim (1972) and Lewis (1969; see also Causey, 1977, p. 149) propose species-specific identities: pain is one brain state in dogs and another in people. As should be clear from this introduction, however, this move sidesteps the main metaphysical question: ''What is common to the pains of dogs and people (and all other pains) in virtue of which they are pains?''

Source II

The second major strand in current functionalism descends from Smart's early article on mind-body identity (1959). Smart worried about the following objection to mind-body identity: So what if pain is a physical state? It can still have a variety of phenomenal *properties*, such as sharpness, and these phenomenal properties may be irreducibly mental. Then Smart and other identity theorists would be stuck with a ''double aspect'' theory: pain is a physical state, but it has both physical and irreducibly mental properties. He attempted to dispel this worry by analyzing mental concepts in a way that did not carry with it any commitment to the mental or physical status of the concepts.[9] These ''topic-neutral analyses,'' as he called them, specified mental states in terms of the stimuli that caused them (and the behavior that they caused, although Smart was less explicit about this). His analysis of first-person sensation avowals were of the form ''There is something going on in me which is like what goes on when . . . ,'' where the dots are filled in by descriptions of typical stimulus situations. In these analyses, Smart broke decisively with behaviorism in insisting that mental states were real things with causal efficacy; Armstrong, Lewis, and others later improved his analyses, making explicit the behavioral effects clauses, and including mental causes and effects. Lewis's formulation, especially, is now very widely accepted among Smart's and Armstrong's adherents (Smart, 1971, also accepts it). In a recent review in the *Australasian Journal of Philosophy*, Alan Reeves declares, ''I think that there is some consensus among Australian materialists that Lewis has provided an exact statement of their viewpoint'' (1978).

Smart used his topic-neutral analyses only to defeat an a priori objection to the identity theory. As far as an argument *for* the identity theory went, he relied on considerations of simplicity. It was absurd, he thought, to suppose that there should be a perfect correlation between mental states and brain states and yet that the states could be nonidentical. (See Kim, 1966; Brandt and Kim, 1967, for an argument against Smart; but see also Block, 1971, 1979; and Causey, 1972, 1977, for arguments against Kim and Brandt.) But Lewis and Smart's Australian allies (notably D. M. Armstrong) went beyond Smart, arguing that something like topic-neutral analyses could be used to argue

for mind-brain identity. In its most persuasive version (Lewis's), the argument for physicalism is that pain can be seen (by conceptual analysis) to be the occupant of causal role *R*; a certain neural state will be found to be the occupant of causal role *R*; thus it follows that pain = that neural state. Functionalism comes in by way of showing that the meaning of 'pain' is the same as a certain definite description that spells out causal role *R*.

Lewis and Armstrong argue from functionalism to the truth of physicalism because they have a ''functional specification'' version of functionalism. Pain is a functionally specified state, perhaps a functionally specified brain state, according to them. Putnam and Fodor argue from functionalism to the falsity of physicalism because they say there are functional states (or functional properties), and that mental states (or properties) are identical to these functional states. No functional state is likely to be a physical state.

The difference between a functional state identity claim and a functional specification claim can be made clearer as follows. Recall that the functional state identity claim can be put thus:

pain = $\%y\mathrm{E}x_1 \ldots \mathrm{E}x_n[T(x_1 \ldots x_n)\ \&\ y \text{ has } x_1]$

where x_1 is the variable that replaced ''pain.'' A functional specification view could be stated as follows:[10]

pain = the $x_1\mathrm{E}x_2 \ldots \mathrm{E}x_n T(x_1 \ldots x_n)$

In terms of the example mentioned earlier, the functional state identity theorist would identify pain with the property one has when one is in a state that is caused by pin pricks and causes loud noises and also something else that causes brow wrinkling. The functional specifier would define pain as *the thing* that is caused by pin pricks and causes loud noises and also something else that causes brow wrinkling.

According to the functional specifier, the thing that has causal role *R* (for example, the thing that is caused by pin pricks and causes something else and so forth) might be a state of one physical type in one case and a state of another physical type in another case. The functional state identity theorist insists that *pain* is not identical to a physical state. What pains have in common in virtue of which they are pains is causal role *R*, not any physical property.

In terms of the carburetor example, functional state identity theorists say that being a carburetor = being what mixes gas and air and sends the mixture to something else, which, in turn ... Functional specifiers say that the carburetor is *the thing* that mixes gas and air and sends the mixture to something else, which, in turn ... What the difference comes to is that the functional specifier says that the carburetor is a type of physical object, though perhaps one type of physical object in a Mercedes and another type of physical object in a Ford. The functional state identity theorist insists that

what it is to be a carburetor is to have a certain functional role, not a certain physical structure.

At this point, it may seem to the reader that the odd disagreement about whether functionalism justifies physicalism or the negation of physicalism owes simply to ambiguities in ''functionalism'' and ''physicalism.'' In particular, it may seem that the functional specification view justifies *token* physicalism (the doctrine that every particular pain is a physical state token), while the functional state identity view justifies the negation of *type* physicalism (the doctrine that *pain* is a type of physical state).

This response oversimplifies matters greatly, however. First, it is textually mistaken, since those functional specifiers who see the distinction between type and token materialism clearly have type materialism in mind. For example, Lewis says, ''A dozen years or so ago, D. M. Armstrong and I (independently) proposed a materialist theory of mind that joins claims of *type-type* psychophysical identity with a behaviorist or functionalist way of characterizing mental states such as pain'' (Lewis, 1980; emphasis added). More important, the functional specification doctrine *commits* its proponents to a functional state identity claim. Since the latter doctrine counts against type physicalism, so does the former. It is easy to see that the functional specification view commits its proponents to a functional state identity claim. According to functional specifiers, it is a conceptual truth that pain is the state with causal role *R*. But then *what it is to be a pain* is to have causal role *R*. Thus the functional specifiers are committed to the view that what pains have in common by virtue of which they are pains is their causal role, rather than their physical nature. (Again, Lewis is fairly clear about this: ''Our view is that the concept of pain ... is the concept of a state that occupies a certain causal role.'')

I suspect that what has gone wrong in the case of *many* functional specifiers is simply failure to appreciate the distinction between type and token for mental states. If pain in Martians is one physical state, pain in humans another, and so on for pain in every pain-feeling organism, then each particular pain is a token of some physical type. This is token physicalism. Perhaps functional specifiers ought to be *construed* as arguing for token physicalism (even though Lewis and others explicitly say they are arguing for type physicalism). I shall give three arguments against such a construal. First, as functional state identity theorists have often pointed out, a *non*physical state could conceivably have a causal role typical of a mental state. In functional specification terms, there might be a creature in which pain is a functionally specified *soul* state. So functionalism opens up the possibility that even if *our* pains are physical, other pains might not be. In the light of this point, it seems that the support that functionalism gives even to token physicalism is equivocal. Second, the *major* arguments for token physicalism involve no functionalism at all (see Davidson, chapter 5, and Fodor, chapter 6). Third, token physicalism is a much weaker doctrine than physicalists have typically wanted.

In sum, functional specifiers *say* that functionalism supports physicalism, but they are committed to a functionalist answer, not a physicalist answer, to the question of what all pains have in common in virtue of which they are pains. And if what all pains have in common in virtue of which they are pains is a functional property, it is very unlikely that pain is coextensive with any physical state. If, on the contrary, functional specifiers have *token* physicalism in mind, functionalism provides at best equivocal support for the doctrine; better support is available elsewhere; and the doctrine is a rather weak form of physicalism to boot.

Lewis's views deserve separate treatment. He insists that pain is a brain state only because he takes ''pain'' to be a nonrigid designator meaning ''the state with such and such causal role.''[11] Thus, in Lewis's view, to say that pain is a brain state should not be seen as saying what all pains have in common in virtue of which they are pains, just as saying that the winning number is 37 does not suggest that 37 is what all winning numbers have in common. Many of Lewis's opponents disagree about the rigidity of ''pain,'' but the dispute is irrelevant to our purposes, since Lewis does take 'having pain' to be rigid, and so he does accept (he tells me) a functional property identity view: having pain = having a state with such and such a typical causal role. I think that most functional state identity theorists would be as willing to rest on the thesis that having pain is a functional property as on the thesis that pain is a functional state.

In conclusion, while there is considerable disagreement among the philosophers whom I have classified as metaphysical functionalists, there is a single insight about the nature of the mind to which they are all committed.

Notes

Reprinted from N. Block, ed., *Readings in Philosophy of Psychology*, vol. 1, 171–184 (Cambridge, MA: Harvard University Press, 1980). Also reprinted in John Heil, ed., *Philosophy of Mind: A Guide and Anthology* (Oxford: Oxford University Press, 2004).

1. Discussions of functional state identity theses have sometimes concentrated on one or another weaker thesis in order to avoid issues about identity conditions on entities such as states or properties (see, for example, Block and Fodor, chapter 20). Consider the following theses:

(1) Pain = functional state *S*.
(2) Something is a pain just in case it is a (token of) *S*.
(3) The conditions under which *x* and *y* are both pains are the same as the conditions under which *x* and *y* are both tokens of *S*.

(1) is a full-blooded functional state identity thesis that entails (2) and (3). Theses of the form of (2) and (3) can be used to state what it is that all pains have in common in virtue of which they are pains.

2. Dennett (1975) and Rey (1979) make this appeal to the Church-Tuning thesis. But if the mechanical processes involved analog rather than digital computation, then the processes could fail

to be algorithmic in the sense required by the Church-Turing thesis. The experiments discussed in Block 1981, part two, ''Imagery'' suggest that mental images are (at least partially) analog representations, and that the computations that operate on images are (at least partially) analog operations.

3. Strictly speaking, even the causal role formulation is insufficiently general, as can be seen by noting that Turing machine functionalism is not a special case of causal role functionalism. Strictly speaking, none of the states of a Turing machine need cause any of the other states. All that is required for a physical system to satisfy a machine table is that the counterfactuals specified by the table are true of it. This can be accomplished by some causal agent outside the machine. Of course, one can always choose to speak of a *different* system, one that includes the causal agent as part of the machine, but that is irrelevant to my point.

4. Formulations of roughly this sort were first advanced by Lewis, 1966, 1970, 1972; Martin, 1966. (See also Harman, 1973; Grice, 1975; Field, 1978; Block, chapter 22.)

5. See Field, 1978, for an alternative convention.

6. Characterizations of mental states along the lines of the Ramsey sentence formulation presented above wear their incompatibility with behaviorism on their sleeves in that they involve explicit quantification over mental states. Both Thomas and Bealer provide ways of transforming functionalist definitions or identifications so as to disguise such transparent incompatibility.

7. The machine table states of a finite automaton can be defined explicitly in terms of inputs and outputs by a Ramsey sentence method, or by the method described in Thomas (1978). Both of these methods involve one or another sort of commitment to the existence of the machine table states.

8. 'Physical state' could be spelled out for these purposes as the state of something's having a first-order property that is expressible by a predicate of a true physical theory. Of course, this analysis requires some means of characterizing physical theory. 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 (but not other properties). The physicalist doctrine that functionalists argue against is the doctrine that mental properties are *first-order* physical properties. Functionalists need not deny that mental properties are second-order physical properties (in various senses of that phrase).

9. As Kim has pointed out (1972), Smart did not need these analyses to avoid ''double aspect'' theories. Rather, a device Smart introduces elsewhere in the same paper will serve the purpose. Smart raises the objection that if afterimages are brain states, then since an afterimage can be orange, the identity theorist would have to conclude that a brain state can be orange. He replies by saying that the identity theorist need only identify the *experience of having an orange afterimage* with a brain state; this state is not orange, and so no orange brain states need exist. Images, says Smart, are not really mental entities; it is experiences of images that are the real mental entities. In a similar manner, Kim notes, the identity theorist can ''bring'' the phenomenal properties into the mental states themselves; for example, the identity theorist can concern himself with states such as John's having a sharp pain; this state is not sharp, and so the identity theorist is not committed

to sharp brain states. This technique does the trick, although of course it commits its perpetrators to the unfortunate doctrine that pains do not exist, or at least that they are not mental entities; rather, it is the havings of sharp pains and the like that are the real mental entities.

10. The functional specification view I give here is a much simplified version of Lewis's formulation (1972).

11. A rigid designator is a singular term that names the same thing in each possible world. 'The color of the sky' is nonrigid, since it names blue in worlds where the sky is blue, and red in worlds where the sky is red. 'Blue' is rigid, since it names blue in all possible worlds, even in worlds where the sky is red.

References

Armstrong, D. M. 1968. *A Materialist Theory of Mind*. London: Routledge & Kegan Paul.

——— 1970. *The Nature of Mind*. In C. V. Borst, ed., *The Mind/Brain Identity Theory*. London: Macmillan.

——— 1977. ''The Causal Theory of the Mind.'' In *Neue Heft für Philosophie*, no. 11, pp. 82–95. Vendenhoek and Ruprecht.

Bealer, G. 1978. ''An Inconsistency in Functionalism.'' *Synthese* 38:333–372.

Block, N. 1971. ''Physicalism and Theoretical Identity.'' Ph.D. dissertation, Harvard University.

——— 1978. ''Troubles with Functionalism.'' In C. W. Savage, ed., *Minnesota Studies in Philosophy of Science*. Vol. 9. Minneapolis: University of Minnesota Press.

——— 1979. ''Reductionism.'' In *Encyclopedia of Bioethics*. New York: Macmillan.

——— 1981. *Readings in Philosophy of Psychology*. Vol. 2. Cambridge MA: Harvard University Press.

Block, N., and J. A. Fodor. 1972. ''What Psychological States Are Not.'' *Philosophical Review* 81, no. 2:159–182.

Brandt, R., and J. Kim. 1967. ''The Logic of the Identity Theory.'' *Journal of Philosophy* 64, no. 17:515–537.

Causey, R. 1972. ''Attribute Identities in Micro-reductions.'' *Journal of Philosophy* 69, no. 14:407–422.

——— 1977. *Unity of Science*. Dordrecht: Reidel.

Chisholm, R. M. 1957. *Perceiving*. Ithaca: Cornell University Press.

Cummins, R. 1975. ''Functional Analysis.'' *Journal of Philosophy* 72, no. 20:741–764.

Dennett, D. 1975. ''Why the Law of Effect Won't Go Away.'' *Journal for the Theory of Social Behavior* 5:169–187.

Field, H. 1978. ''Mental Representation.'' *Erkenntniss* 13:9–61.

Fodor, J. A. 1965. ''Explanations in Psychology.'' In M. Black, ed., *Philosophy in America*. London: Routledge & Kegan Paul.

——— 1968a. ''The Appeal to Tacit Knowledge in Psychological Explanation.'' *Journal of Philosophy* 65:627–640.

——— 1968b. *Psychological Explanation*. New York: Random House.

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

Gendron, B. 1971. ''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*. Vol. 8. Dordrecht: Reidel.

Grice, H. P. 1975. ''Method in Philosophical Psychology (from the Banal to the Bizarre).'' *Proceedings and Addresses of the American Philosophical Association*. Newark, Del.: American Philosophical Association.

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

Hartman, E. 1977. *Substance, Body and Soul*. Princeton: Princeton University Press.

Kalke, W. 1969. ''What Is Wrong with Fodor and Putnam's Functionalism?'' *Nous* 3:83–93.

Kim, J. 1966. ''On the Psycho-physical Identity Theory.'' *American Philosophical Quarterly* 3, no. 3:227–235.

——— 1972. ''Phenomenal Properties, Psychophysical Law, and the Identity Theory.'' *Monist* 56, no. 2:177–192.

Lewis, D. 1966. ''An Argument for the Identity Theory.'' Reprinted in D. Rosenthal, ed., *Materialism and the Mind-Body Problem*. Englewood Cliffs, N.J.: Prentice-Hall, 1971.

——— 1969. ''Review of *Art, Mind and Religion*.'' *Journal of Philosophy* 66, no. 1:23–35.

——— 1970. ''How to Define Theoretical Terms.'' *Journal of Philosophy* 67, no. 13:427–444.

——— 1972. ''Psychophysical and Theoretical Identification.'' *Australasian Journal of Philosophy* 50, no. 3:249–258.

——— 1980. ''Mad Pain and Martian Pain.'' In N. Block, ed., *Readings in Philosophy of Psychology*, Vol. 1, 216–222. Cambridge, MA: Harvard University Press.

Lycan, W. 1979. ''A New Lilliputian Argument against Machine Functionalism.'' *Philosophical Studies* 35, 279–287.

Martin, R. M. 1966. ''On Theoretical Constants and Ramsey Constants.'' *Philosophy of Science* 31:1–13.

Nagel, T. 1970. ''Armstrong on the Mind.'' *Philosophical Review* 79:394–403.

Nelson, R. J. 1975. ''Behaviorism, Finite Automata and Stimulus Response Theory.'' *Theory and Decision* 6:249–267.

——— 1976. ''Mechanism, Functionalism and the Identity Theory.'' *Journal of Philosophy* 73, no. 13:365–386.

Putnam, H. 1960. ''Minds and Machines.'' In S. Hook, ed., *Dimensions of Mind*. New York: New York University Press.

——— 1963. ''Brains and Behavior.'' Reprinted in *Mind, Language, and Reality: Philosophical Papers*. Vol. 2. London: Cambridge University Press, 1975.

——— 1966. ''The Mental Life of Some Machines.'' Reprinted in *Mind, Language and Reality: Philosophical Papers*. Vol. 2. London: Cambridge University Press, 1975.

——— 1967. ''The Nature of Mental States'' (originally published as ''Psychological Predicates''). In W. H. Capitan and D. D. Merrill, eds., *Art, Mind, and Religion*. Pittsburgh: University of Pittsburgh Press.

——— 1970. ''On Properties.'' In *Mathematics, Matter and Method: Philosophical Papers*. Vol. 1. London: Cambridge University Press.

——— 1975. ''Philosophy and Our Mental Life.'' In *Mind, Language and Reality: Philosophical Papers*. Vol. 2. London: Cambridge University Press.

Reeves, A. 1978. ''Review of W. Matson, *Sentience*.'' *Australasian Journal of Philosophy* 56, no. 2 (August):189–192.

Rey, G. 1979. ''Functionalism and the Emotions.'' In A. Rorty, ed., *Explaining Emotions*. Berkeley and Los Angeles: University of California Press.

Ryle, G. 1949. *The Concept of Mind*. London: Hutchinson.

Sellars, W. 1968. *Science and Metaphysics*. London: Routledge & Kegan Paul, chap. 6.

Shoemaker, S. 1975. ''Functionalism and Qualia.'' *Philosophical Studies* 27:271–315.

Smart, J. J. C. 1959. ''Sensations and Brain Processes.'' *Philosophical Review* 68:141–156.

——— 1971. ''Reports of Immediate Experience.'' *Synthese* 22:346–359.

Thomas, S. 1978. *The Formal Mechanics of Mind*. Ithaca: Cornell University Press.