

This is a preprint of an article whose final and definitive form will be published in the *Routledge Companion to the Philosophy of Psychology*

<http://www.routledge.com>

Computational Functionalism

Thomas W. Polger

Associate Professor

Department of Philosophy

University of Cincinnati

Cincinnati, Ohio 45221-0374 USA

thomas.polger@uc.edu

<http://homepages.uc.edu/~polgertw>

1. Computational Functionalism

To a first approximation, computational functionalism is the view that the mind is usefully thought of as a mechanical device, such as a computing machine. The core idea behind the mind-as-machine theory is that psychological entities should be understood in terms of what they do rather than in terms of the stuff of which they are made. Thus computational functionalism departs from both traditional materialism and dualism, and exhibits some affinity with the behaviorist approach.

The core idea of computational functionalism has been developed in various ways so that functionalism now constitutes a whole family of theories. Functionalist theories vary along several dimensions, including (a) how broadly the theories apply, (b) how they unpack the notion of a computing device or machine, and (c) whether they take the mind-machine comparison literally or metaphorically. But functionalist theories also share some core commitments, and

they are generally advanced on the basis of a few common arguments. In what follows we will explore these differences and commonalities.

Before moving forward, two caveats are in order. First, computational functionalism should not be confused with other “functionalist” theories in psychology and philosophy of science. In particular, and despite some similarities, contemporary computational functionalism is distinct from the “Chicago” functionalism advocated by American pragmatist psychologists in the late 19th and early 20th centuries.¹ Henceforth I will drop the qualification “computational” and speak generically of “functionalism” except where some confusion might arise from this terminological economy. Second, I cast my net widely when collecting together the family of functionalist theories. Some philosophers and psychologists reserve the term “functionalism” for only a subset of the variations to be discussed herein. For example, some philosophers take functionalism only to be a thesis about the narrow or internal relations among psychological states; on something like this basis, Fred Dretske (1995) denies that he is a functionalist. And Ned Block (1980) uses “Functionalism” (with a capital ‘F’) to refer to the particular version of the theory that is exemplified by the work of David Lewis (1966, 1972).² I trust that my less restricted use of the term will be vindicated if it allows us to systematize and explain a range of philosophical views. So let us now proceed to doing so.

2. Varieties of Functionalism

Begin with Brentano’s Thesis, the idea that “intentionality is the mark of the mental.” On the view associated with Brentano, to be a psychological (i.e., mental) entity is to be intentional, to

¹ Space limitations prevent us from taking up herein the interesting question of the origins and history of functionalism. For some preliminary thoughts on the matter, see my 2004.

² Lewis himself called this an “identity theory” early on (1966, 1972), and later asserted that it was a reductionist theory while deferring the question as to whether it is a functionalist theory (1994).

have content. Take paradigmatic intentional states such as beliefs and desires. Plainly beliefs mediate our perceptions and actions differently than desires. My belief that it is a sunny day is usually caused by seeing or being told that the day is sunny but my desire that it be a sunny day is not caused in that way, for example. And beliefs mediate our perceptions and actions differentially depending on their content. My beliefs about sunny days have different causes and effects than my beliefs about coffee. And my beliefs about sunny days are related to other beliefs — such as those about blue skies, about beaches, and so forth — which are different than those to which my beliefs about coffee are related.

Suppose that some physical state of a system mediates the perceptual inputs and behavioral outputs of the system in the ways that are characteristic of a belief, and that it is related to objects in the world and to other states of the system in the ways that are distinctive of a belief with a particular propositional content, say, that there is coffee in this mug. If so, then why should we not say that the physical state in question simply is the belief that there is coffee in this mug? If it looks like a duck and quacks like a duck, then it's a duck. So the platitude goes. The core functionalist idea is that a state that acts (all-things-considered) like a belief or a pain simply is a belief or a pain, respectively.

There are a number of attractive features of the functionalist view, not least of which is that it may seem like plain common sense. Second, functionalism makes the conditions for being or having a mental state quite independent of the material substance (or immaterial substance, for that matter) of which a system is composed. This seems to have the corollaries that there could be non-biological mental entities — naturally occurring aliens or artificially engineered robots — and that the study of psychology may proceed more or less independently of the study of the physical (or non-physical) systems in which psychology is exhibited. Third, functionalism

seems to bring the conditions for being a mental entity directly in line with ordinary and empirical criteria for determining what mental states a system has. Functionalists deny that anything could play the full role of a belief about coffee without thereby being such a belief. So functionalism provides a framework for psychological explanation, a methodological model for studying psychology, and a theoretically motivated response to skeptics about other minds.

All this is to say that, as illustrated above, functionalism promises to answer a bevy of philosophical questions about minds in one blow: We get a theory (or set of related theories) about (1) the *metaphysics* or ontology of mental states, (2) the *explanatory* and (3) the *theoretical* structures of psychological science, (4) the *intentionality* and (5) the *semantics* of mental states, and (6) the *methodology* for the study of psychology. As these different phenomena run the full range of those that have concerned philosophers of mind and psychology, functionalism promises to be a comprehensive theory.³

It is unclear whether any single theorist has ever endorsed the functionalist theory for all six of these phenomena. Determining the answer is troublesome because it is often difficult to discern when functionalism is put forth as a serious theory and when it is advanced as a metaphor or a convenient idealization. But each variety of functionalism can be found separately among contemporary theorists.

2.1. Metaphysical functionalism

Begin with functionalism as a metaphysical theory about the nature of psychological states or processes. The slogan “mental states are functional states” is often used to express the functionalist metaphysical thesis. Here the idea is that some states or properties are such that

³ On the range of problems in philosophy of mind, see Paul Churchland (1988) from whom I adapt this list. The discussion in what follows relies heavily on the development of this taxonomy in chapter 3 of my 2004, where the characterization of each variety is developed more thoroughly.

they are wholly constituted by their relations to one another and to certain inputs and outputs, and that mental states or properties are among these. For example, what makes something (a rock, say) into a doorstop is not its inner structure or essence (if any), but simply that it is suitable for playing the role of a doorstop — that it functions as a doorstop. Being a doorstop is wholly a matter of a thing's relations to certain other objects, especially floors and doors. Likewise, what makes something a calculator is that it functions as a calculator, regardless of the stuff of which it is made.

In the jargon of functionalism, doorstops and calculators are *realized* by various physical systems, but neither doorstops nor calculators are physical natural kinds. Their various potential realizers are a physically heterogeneous group, having in common only that they realize (equivalently: implement, function as, or play the role of) doorstops and calculators, respectively. The clearest examples of such “functional” kinds are artifacts. But a case can be made that biological kinds, such as hearts or eyes, are functional kinds in the same sense.

Metaphysical functionalism is the theory that the ontological nature of mental states is that they are realized by physical states of brains. Thus mental states are not identical to brain states, as the psycho-physical identity theory holds; nor are mental states simply syndromes of behavior and behavioral dispositions, as some behaviorists held. Rather mental states are the functionally constituted internal states, e.g., brain states or machine states, whose symptoms are various bits of behavior and behavioral dispositions of their containing systems. Functional states are understood to be relational in nature. Being a mental state is thus like being on the left of something rather than like having a mass of five kilograms; it essentially involves a thing's relations to other things, not merely the properties of the thing in itself. The functionalist holds that a realizer of a psychological state must be a thing of some sort — for it must be able to enter

into the correct functional relations — but it need not be a thing of any one particular sort. In this way, functionalism aims to plot a middle course between the “chauvinistic” excesses of the identity theory — which seems to imply that only things with brains relevantly like ours can have mental states — and the “liberal” excesses of behaviorism — which counts any apparently intelligent behavior as genuinely psychological regardless of how it is produced.⁴

Metaphysical functionalism does not merely claim that psychological states are like functional states or can be usefully thought of as such. Metaphysical functionalism takes the functionalist slogan that “mental states are functional states” to express a literal truth. Hilary Putnam, in his “The Nature of Mental States,” is usually credited as the first functionalist, for advancing metaphysical functionalism as an empirical hypothesis (1967).⁵ Putnam characterized functional states in terms of probabilistic automata, a generalization of Alan Turing’s finite state machines. He and others had for years thought of these computing devices as good models for human psychology. Indeed Turing himself had set forth a behavioral test for intelligence that he was sure such machines would be able to pass.⁶ But it was Putnam who upgraded the idea from a metaphor or model and articulated it as an ontologically committed theory of the nature of mental states.

While the functionalist slogan that “mental states are functional states” can be used to express the thesis of metaphysical functionalism, it can also be used to express a variety of ontologically weaker claims. It is to these that we now turn.

⁴ The “liberal” versus “chauvinistic” distinction comes from Block (1978). Obviously the simplified representation of both identity theories and behaviorist theories is a caricature, but it is a useful one for seeing the general motivation behind functionalism.

⁵ There is an interesting question about whether various earlier philosophers — from Aristotle to Sellars — held a view that we would call functionalist. But Putnam appears to be the first to explicitly advocate the view, and he is the one who gave it its name and distinctive formulation in terms of computing machines.

⁶ Turing (1950) expected the machines to pass this test by the year 2000. To date, no machine has come close to satisfying the requirements of a generalized “Turing Test.” Turing appears to have underestimated the technical challenges no less than the philosophical.

2.2. Explanatory functionalism

Rather than taking the view that mental states are functional states to be an ontological thesis about the nature of psychological entities, one might take functionalism as an explanatory framework that allows us to remain neutral about the ontological status of psychological entities. This is one of the applications of explanatory functionalism.

I take it that explanatory functionalism is something like the view that a phenomenon x is best explained by reference to its function, or to the functioning x 's constituents to enable its behavior. On this general formulation, every functionalist theory is a case of explanatory functionalism, where the variable x may be replaced by the ontology of mind, the intentionality of beliefs, and so on. Thus explanatory functionalism turns out to be the central kind from which all others can be formulated. There is something to this idea. But those who are realists about the phenomena in question — and functionalists have tended to be realists — will think that explanations of psychological ontology, intentional content, and so forth take the functional form precisely because those phenomena are essentially functional in themselves. So amenability to functional explanation does not constitute the truth of functionalisms, rather it is a sign of the common functional natures of their explananda.

For the purposes of the present taxonomy we are interested in a stronger form of explanatory functionalism, one that makes specific commitments about the explanatory structure of psychology. Classical behaviorism was the view that psychology is the science of behavior and psychological explanations should be stimulus-response explanations. Likewise, the explanatory versions of functionalism hold that psychological explanations should be functional explanations — explanations in terms of inputs, outputs, and relations among internal states.

This means that one has not completed a psychological explanation merely by describing a lawful or lawlike regularity between inputs and outputs; in addition, psychological explanation requires a characterization of the mechanism that produces the outputs (and changes in other internal states) in response to the inputs. This kind of explanatory activity is what Robert Cummins (1975) calls “functional analysis” and Daniel Dennett (1981) calls “homuncular decomposition.” A functional explanation succeeds in explaining how a system operates by showing how its component parts contribute to the operation (behavior) of the total system through their interactions with one another and with inputs and outputs. The best known account of functional explanation in psychology is due to David Marr, who describes how layers of functional explanation interact in a theory of human vision (1982).

The advantage of functional explanation is that it can proceed independently of most knowledge of the internal structure, if any, of the functional components. This independence is possible because the explanation appeals only to the relations of those components with one another and the inputs and outputs of the containing system. As a consequence, it is widely held, functional explanations capture regularities that are either invisible or practically intractable to explanations in terms of components’ structures or substantial composition. To take a simple example, one can explain the operation of a vending machine by appeal to how the system responds to various inputs, e.g., coins, to produce certain outputs, e.g., bottles (e.g., Block 1980). The functional explanation will look very much like a program or decision procedure. And it depends only on the functional relations among the states of the machine and not on the particular physical processes in which the system is implemented. Explaining the vending behavior in terms of the laws of physics would be difficult or impossible, but it is an easy task for functional explanation (Dennett 1971, Putnam 1975b).

The net result is that functional explanations are often thought to be *autonomous* from the explanations of, say, physics. A functional explanation may be compatible with any of a variety of physical explanations of an explananda, but its value does not await delivery of the physics explanation or depend on the explanatory terms or entities have any special relation to one another. The view that functional explanations are autonomous stands in contrast with classical reduction, according to which the “higher level” explanatory terms are required to be coextensive with those of the “lower-level” explanation, so that the natural kinds of the “higher-level” explanation are also natural kinds of the “lower-level” explanation. Jerry Fodor (1974) and Philip Kitcher (1982) persuaded many philosophers that scientific and explanatory legitimacy does not require coextension of predicates. So if psychological explanations are functional explanations, that seems to show how they can eschew such constraints. Psychology, then, is vindicated as a science in its own right.

2.3. Theoretical functionalism

If psychological explanations are functional explanations, and if theories are sets of explanations, then there is a simple way in which psychological theory would turn out to be a functional theory: It would be constituted by functional explanations. This would be an interesting and important result. It is by no means obvious that actual scientific psychology is a functional theory of even this weak sort.

Yet there is a more demanding notion of functional theories that is widely applied in philosophy of psychology. This stronger idea of a functional theory begins with a psychological theory that is functional in the weak way, and then uses such a theory to define the theoretical terms of psychology (Lewis 1970, 1972). The resulting theory is a functional theory in that the

theoretical terms of the science can be wholly defined by the role that they play in the total theory. Suppose, for example, that our psychological explanations include the following sorts of claims: For all creatures of kind K, when they are exposed to acute skin damage or...etc. they experience pain, which causes them to form a belief that the stimulus is harmful and causes them to attempt to withdraw the affected area from the stimulus or remove the stimulus or...etc. Of course there would be many such explanatory schema or laws regarding pain, regarding memory, regarding belief, and so forth. The proposal of the theoretical functionalist is that we can conjoin the explanatory regularities of a completed psychology together and use them to define the meanings of its theoretical terms. So we can say the meaning of the term 'pain' is: "the state such that, for all creatures of kind K, when they are exposed to acute skin damage or...etc. they experience pain, which causes them to form a belief that the stimulus is harmful and causes them to attempt to withdraw the affected area from the stimulus or remove the stimulus or...etc."⁷

One might adopt theoretical functionalism but still hold that the functional theory does not explain the real natures of psychological states because it describes only their relational properties and none of their intrinsic properties. A substance dualist could think that theoretical functionalism is the right semantic account for the theoretical terms of psychology. But it will at least be tempting to hold that a term-defining functionalist theory implicitly says everything that there is to say about the entities to which the theoretical terms of psychology apply. After all, it will be a definitional or analytic truth that pain is the state such that, for all creatures of kind K, when they are exposed to acute skin damage or...etc. they experience pain. And likewise that the pain causes creatures of kind K to form a belief that the stimulus is harmful and causes them to attempt to withdraw the affected area from the stimulus or remove the stimulus or...etc. So

⁷ Some philosophers believe that every term that does not pick out a logical relation or a basic element of our ontology (the ontology of physics, if we are physicalists) must be defined in this way (Lewis 1994, Chalmers 1996, Jackson 1998.)

while theoretical functionalism does not strictly entail metaphysical functionalism, it provides one natural route to the metaphysical view.

2.4. Intentional functionalism

Thus far I have focused on functionalism as account of what it is to be a psychological state, and as an account psychological explanation and the content and structure of psychological theories.

But two important varieties of functionalism offer accounts of the contents of psychological states themselves. The first, which I call *intentional functionalism*, is the view that what makes some states into intentional states — states that have content, aboutness — is their functional role. I gave an example of this view earlier: One might think that beliefs have a distinctive role in modulating behavior, perhaps to be understood in terms of the characteristic ways that beliefs are caused by sensory stimulation and (when interacting with appropriate desires) result in characteristic actions. On this view, what makes some state an intentional state at all, and moreover the kind of intentional state that it is (a belief in the above example) is that it plays the belief-role: As Fodor explains the idea:

to hold that to believe it's raining is to have a token of 'it's raining' play a certain functional role in the causation of your behavior and of your (other) mental states, said role eventually to be specified in the course of the detailed working out of empirical psychology..., etc., etc. This is, perhaps not much of story, but it's fashionable, I know of nothing better, and it does have the virtue of explaining why propositional attitudes are opaque. (Fodor 1978 in Block 1980: 53-54)

The difference, then, between distinct kinds of intentional states, e.g., beliefs and desires, will be a difference in their characteristic roles. Likewise, what distinguishes beliefs from non-

intentional states that mediate stimuli and behaviors will be explained in terms of their distinctive role. States of the spinal nerves that mediate reflexive withdrawal from painful stimuli may be distinguished from beliefs by their functional role — for example, their general failure to interact with other beliefs and with desires, etc.

Of course, if one holds that what it is to be an intentional state is a matter of functional role and additionally holds Brentano's Thesis that intentionality is the mark of the mental, then it follows that what it is to be a mental state is itself a matter of functional role. Thus intentional functionalism and Brentano's Thesis entail metaphysical functionalism about psychological states.

2.5. Semantic functionalism

The second variety of functionalism about content is semantic functionalism. Semantic functionalism is the view that the particular content of a mental state is constituted by its functional role. I distinguish intentional functionalism from semantic functionalism. The first is a view about what makes some bits of the world into content bearing states at all, and of which kinds. The second is a view about what gives such intentional states their particular content. In the example I used earlier, semantic functionalism is illustrated by the idea that what makes my beliefs about sunny days distinct from my beliefs about coffee is that the different beliefs play different roles — one group typically mediates my interactions with weather, and the other with beverages.

It is attractive to think that the intentional and semantic questions can be answered at the same time, and even by the same relations: Perhaps what makes something a belief about trees is that it plays the belief role relative to stimulation from and behavior toward trees, as well as

relative to beliefs about leaves and bark and desires about shade and apples. And, indeed, some philosophers espouse both functionalism about intentionality and functionalism as a theory of psychosemantics, that is, of the semantic content of psychological states.

But it is also possible, and indeed common, to hold only one of the two functionalist theories. For example, one could have the idea that what gives a belief its specific content is the role that belief plays in a network of other mental states without paying any attention to the question of what makes something a belief in general. This would be an ordinary version of functional role semantics for mental states (cf. Van Gulick 1980). And, on the other hand, some prominent theorists who are functionalists about intentionality, e.g., William Lycan (1987) and Jerry Fodor (1986), reject functionalism and general account of psychosemantics. So whereas it is possible to hold both intentional and semantic functionalisms together, it is certainly not mandatory.⁸

2.6. Methodological functionalism

Like explanatory functionalism, methodological functionalism has generic and robust instances. Any method of studying a system by observing and characterizing its inputs and outputs will be functionalist in the generic sense. It is this generic usage that sometimes tempts philosophical supporters of functionalism to think it is ubiquitous and critics to think it is trivial. A method of psychological inquiry or analysis is robustly functional if it treats its objects as functional in one of ways discussed above. Two such methods are common. The first is the method of functional analysis or decomposition, mentioned with respect to explanatory functionalism, in section 2.2

⁸ Here I leave the functionalist theory of the semantics of words aside, but for the discussion of functional accounts of theoretical terms, above. But functionalism about linguistic meaning is clearly closely related to functionalism as a view of psychosemantics. Both the psychological and linguistic versions are influenced by Wittgenstein's doctrine of meaning as use (1953).

above. The second is the method of functionally defining theoretical terms, mentioned with respect to theoretical functionalism, in section 2.3 above.

Any activity that results in functional explanations is ipso facto a functionalist methodology in a way. This makes functionalist methods abundant. But this is also a very minimal claim. It amounts to little more than saying that functionalist methods are those that study how things function. Psychology is obviously functional in this sense — regardless of whether it is behaviorist, cognitive, or biological psychology we have in mind. But the claim is correspondingly weak and uninteresting. The more pressing questions concern whether psychology is functionalist in some more robust way. In particular, philosophers have been interested in whether psychology is metaphysically, intentionally, or semantically functionalist, and in whether the theoretical terms of psychology are susceptible to definition in the way hypothesized by theoretical functionalism.

But we must be cautious in evaluating the candidates. Take for example the methods of functional neuroimaging, including functional magnetic resonance imaging (fMRI). Such techniques have rightly received much attention of late. But is functional MRI a method that treats its object of study as essentially functional in nature? This is much less clear. Imaging studies try to give us a picture — literally and theoretically — of what is going on in subjects' brains when they perform various activities, e.g., remembering words, viewing visual stimuli, listening to auditory cues, etc. But none of this makes any assumption that the brain processes observed and the mental processes are essentially functional. While fMRI and similar techniques certainly study brains while they are functioning — in action, as it were — there is nothing intrinsic to the techniques or the resulting theories that requires a commitment to any

distinctively functionalist thesis. Of course some theorists may propose functionalist hypotheses about neural and psychological functions, but that is another matter.

3. Functionalism, Function, and Realization

Thus far I have neglected two crucial questions about functionalism: What is a function? And what does it take to realize a function or play a functional role?

The most common notion of function employed in philosophy of mind is a causal notion of function. The function of a thing is characterized by its causes and effects, by what it does or is apt to do. It follows that to realize that sort of function or have that kind of functional role is to be a thing that has those characteristic causes and effects (e.g., Kim 1998, Shoemaker 2001, 2003, Melnyk 2003, Gillett 2002, 2003). My earlier examples of hearts and doorstops implicitly appealed to this idea: Hearts are things that pump blood; doorstops are things that hold open doors. I also employed this idea when I discussed the varieties of psychological function; for example I talked about beliefs mediating causal interactions with the world in characteristic ways, ways that differ from those in which desires mediate behavior.

But there are also other notions of function that have been employed in functionalist theories of psychology. Early proposals seem to have had in mind some kinds of abstract functional relations from mathematics, from semantics, or from information theory. If the salient functional relations are abstract like the relations among numbers, then realizing those in a physical system is usually held to be a matter of having physical states that are isomorphic to or map onto the abstract relations (Van Gulick 1988, Cummins 1989). And recently some philosophers have appealed to teleological and etiological notions of function drawn from

evolutionary biology (Millikan 1984, Lycan 1987). According to these views, the function of thing is (roughly speaking) the effect that its ancestors had for which they were selected by natural selection. For example, contrary to what was suggested earlier, it might be that being a heart requires neither actually pumping blood nor being apt to do so. Instead being a heart requires being a thing of a kind that conferred an evolutionary benefit to its ancestral bearers because things of that kind in fact pumped blood in them. It follows that to realize this sort of function requires having a certain evolutionary history, rather than having some causal powers or standing in any mapping relation. One has to have had the right ancestors, and so forth.

There are no doubt other notions of function that could be explored, and hybrid variations as well (e.g., Lycan 1987, Dretske 1995). And there is no reason that any notion of function could not be tried in any variety of functionalism. Likewise there is no reason to suppose that functionalism is an all-or-nothing affair. One might be a functionalist about mental content but not about psychological ontology. One might apply the etiological account of function to psychosemantics and the causal account to one's psychological explanations.

As I've characterized functionalism, it may be formulated in terms of various kinds of functions, each of which carries its own conditions for realization. So realization is a matter of having a function of some sort (Polger 2004, forthcoming). But mine is not the only approach to understanding the realization relation; and its articulation is at least as controversial as that of functionalism (see, e.g., Poland 1994; Kim 1998, 2005; Wilson 2001; Shoemaker 2001, 2003; Gillett 2002, 2003; and Melnyk 2003).

Assessing the merits and demerits of all the varieties of functionalism is another matter; at present we are only concerned with identifying the variants. It seems obvious to me that an adequate account of functionalism ought to leave room for all of these variations, and that an

adequate account of realization ought not to assume that only one kind of function can be realized. But I have already noted that my equanimity is controversial.

4. From Metaphor to Theory: Arguments for Functionalism

The slogan that psychological states are functional states is, on one reading, obvious.

Psychological states are states that have functions in thinking. Psychological states are states that mediate between inputs and outputs of a system, and among other internal states of the system some of which may also be psychological states. This is just to say that psychological states do various things; but that observation is not very informative. It is certainly not a philosophical theory about psychology, psychological states, or their contents.

Even the somewhat more substantial thesis that psychological states are computational states leaves too many important details unspecified. It has been widely noted that on at least some conceptions of computation, every state of every object can be treated as a computational state (Putnam 1988, Searle 1990). Even if that conclusion does not follow from every account of computation, it at least shows that the computational functionalist owes us an explanation of his or her idea of computational states before we know what to make of the claim that psychological states are states of that sort.

But the deeper problem is that the mere claim that psychological states are functional states has a simple predicative use, saying of psychological states that they have a certain property, viz., a certain function. This account falls short of the claim that the property in question — being a functional state, as of a computing machine — is essential or even important to or explanatory of the state's being psychological, intentional, semantic, and so on.

A full blown functionalist theory goes beyond noticing certain commonalities between psychological states and functional states, it offers as an explanation of those commonalities the suggestion that mentality (or intentionality, or psychological explanation, etc.) is essentially functional in something like the ways discussed in section 2. There are many reasons for making the upgrade from the functional view of the mind as a metaphor to a fully formed functionalist theory of mind. I will focus on the three that have been most influential.

Probably the most well-known reason to adopt a functionalist theory is given by the *multiple realization* argument originated by Putnam (1967). He argues that the best explanation for the prevalence of psychological states in nature given the heterogeneity of psychological systems is that mental states are internal states of creatures (contra behaviorism) but that they should be identified with functional states of whole organisms rather than relatively local states of brains (contra the mind-brain type identity theory.) This proposal accounts for the apparent fact that the same psychological state, e.g., pain, can be had by creatures that are quite different anatomically and neurologically, e.g., human beings, dogs, and octopi. That is, the fact that psychological states are realized by multiple physically diverse systems. Although the fact of multiple realization does not entail the truth of functionalism, it is highly suggestive. For it is widely thought that if the multiple realizability argument succeeds, then the identity theory is false. With the decline of behaviorism in psychology and philosophical psychology, the identity theory and functionalism are the main standing alternatives for a theory of the nature of metaphysics of psychological states and processes. So most philosophers regard Putnam's argument as more or less decisive against the type identity theory. In this context the multiple realizability argument can be treated as the most widely accepted argument for functionalism.

The second prominent argument in favor of functionalism in psychology is the autonomy argument articulated by Putnam (1975b) and Jerry Fodor (1974). According to Putnam and Fodor, the so-called special sciences (viz., the sciences other than physics and maybe chemistry) typically get their explanatory power by characterizing regularities with a kind of generality that abstracts away from the physical details of systems. Psychology is a paradigm special science in this sense. Take any example of folk psychological explanation in terms of belief and desire or of scientific psychological explanation in terms of information and representation. Belief, desire, information, and representation are all characterized in ways that do not even mention the physical properties of the systems that bear them.⁹ (Because they are “topic neutral” the objects of these characterizations are also open to multiple realization, a connection not lost on Putnam and Fodor.) This seems to show that psychology should go about its business without much need to know about neurophysiology or whatever other “hardware” the psychological systems are implemented in. Psychology is thus vindicated as an autonomous and independent science.

But the autonomy of psychology is only one side of the coin. The other side is that since cognitive psychologists were already operating more or less independently of neuroscientists and because they were already inspired by computational models and metaphors, functionalism seems to provide the best description of the actual practices of psychologists. Work by Miller (1956), Chomsky (1965), Shepard and Metzler (1971) were early influences; but the epitome of this apparent convergence is Marr’s computational theory of vision (1982). The resemblance is clear enough in the cases of explanatory and methodological functionalism, and at least suggestive in the case of metaphysical functionalism. Thus the convergence of philosophical functionalism with the methods and practices of scientific psychology constitute a third line of

⁹ Not coincidentally, they all are characterized intentionally (see, e.g., Dennett 1971.)

reasoning in support of the conclusion that functionalism is the correct philosophical theory of the mind. So to a philosopher of science it may seem obvious that both prescriptive and descriptive considerations support some form of functionalism about psychological states and psychological science.

5. Consciousness, Causal Power, and Constraints: Arguments Against Functionalism

Functionalism, however, is not without its critics. The best known family of objections center around the apparent failure of conscious experiences and sensations to yield to functional analysis or definition (e.g., Block 1978, Levine 1983). It seems as though two systems could be functionally identical (in whatever sense of function one chooses) and yet differ with respect to their states of consciousness. Most famously, we seem to be able to imagine creatures who are functionally indistinguishable from us but whose conscious states are inverted with respect to our own (e.g., Shoemaker 1982, Block 1990), or who lack consciousness altogether (Kirk 1974, Chalmers 1996, Polger 2000). If such scenarios can be sustained, then functionalism is not the correct theory of psychology in general, for there are at least some psychological differences — viz., differences in conscious experience — that are not accounted for by functional differences.

A second important class of objections to functionalism raises problems about the causal efficacy of functional states. The puzzle can be put intuitively using the example of the rock that realizes a doorstep. The rock is an object with a certain size, mass, and so forth. Presumably these physical qualities determine (or are determined by) the causal powers of the rock. Its having these powers is what makes it suitable to be a doorstep. But of course it has those powers whether or not it is playing the role of a doorstep; so the functional property of being a doorstep

doesn't add any new causal powers to the rock. The functional property of being a doorstep seems to be epiphenomenal. We might try saying that the powers of the doorstep are simply the same as those of the rock; but then we would have two objects (a rock and a doorstep) with the same causal powers. So the powers of the thing qua doorstep are either none at all or else simply redundant of those of the rock.¹⁰ Jaegwon Kim, in particular, has pressed this "causal exclusion" problem (1993, 1998, 2005).

Finally, recent critics of functionalism have focused directly on the alleged multiple realization of psychological states. If psychological states are multiply realized, then this is good news for functionalism. But if functionalists overestimate the degree of multiple realization, the theory will turn out to be overly liberal — assigning mental states to too many systems. For example, William Bechtel and Jennifer Mundale have argued that the practices of neuroscientists presuppose that psychological states are not multiply realized (1999), and Lawrence Shapiro has argued that the hypothesis that there are substantial physiological constraints on psychological states is biologically more probable given the evidence we have thus far (2000, 2004).¹¹

So despite being the most widely held cluster of theories in philosophy of psychology, there are some serious challenges that the advocates of functionalism must face. This is hardly the curtain call for functionalism, however. After all, the competing theories all face hard questions as well. It should be no surprise that theories in psychology and philosophical psychology will not generally stand or fall based on philosophical arguments alone. Since functionalism is such a theory, we should expect that its evaluation will be a long-term and multidisciplinary project.

¹⁰ The problem only gets worse when we consider abstract or etiological notions of function, rather than the causal notion appealed to in this example.

¹¹ For a general discussion of the state of play, see Polger 2000 (expanded in 2004) and Bickle 2006.

6. Functionalism, Reduction, and the Autonomy of Psychology

Functionalism, as noted in section 4, is widely thought to be the doctrine about psychological entities that is best able to vindicate psychology as the science of the mind. But this apparent consensus masks a deep divide over exactly how the result is achieved.

On one hand, some philosophers worry that the legitimacy of psychology requires bringing its entities and methods into the general framework of empirical sciences. In particular, they think that it needs to be shown that mental entities and processes are ultimately physical in nature, rather than being distinctively psychical and *sui generis*. For these philosophers, functionalism succeeds because it allows us to explain how physical states and processes can realize or implement psychological states and processes (Putnam 1967, Kim 1998). Thus psychological states are in one way “nothing over and above” physical states, and psychological entities may be (in an important sense of the term) “reduced” to physical entities.

On the other hand, some philosophers worry that the legitimacy of psychology is threatened not only by the possibility that psychological states are essentially non-physical, but also by the possibility — touted in the previous paragraph — that psychological states are nothing more than physical states. For they fear that if psychological states are “nothing over and above” physical states then one might do away with psychology altogether and simply go about the study of the physical realizers.¹² To these philosophers, the success of functionalism is that it shows how mental states can be in some sense physical states, *viz.*, physically realized states, while also showing that they are not identical with and thus not “reducible” to merely physical states (e.g., Fodor 1974, and Kitcher 1982). On this view, psychology is an independent

¹² And this fear is not mere paranoia, for the tactic has been endorsed by various philosophers (P. M. Churchland 1982, P. S. Churchland 1983, Bickle 1996, 1998).

and irreducible science, and the entities that it studies are ontologically real and explanatorily legitimate in their own right.

The lesson is that functionalism has an uneasy relationship with ontological and explanatory reductionism and the complementary autonomy or independence theses. The very consequences that some advocates tout are those that other advocates try to mitigate. While some theorists have questioned whether the connection between functionalism, reduction and autonomy runs as deep as the above sketches suppose (Richardson 1979, Polger 2004), there is no doubt that one of the main attraction of functionalism continues to be its promise to strike a balance between the reduction and autonomy of psychology.

7. Conclusion

I do not claim that my way of thinking about functionalism is the only one, but I've tried to provide one useful way of categorizing its varieties. Even more briefly, I've sketched a few of the most common lines of reasoning for and against functionalism. The sheer number of its varieties assures that there will be continued interest in the project of evaluating functionalism. I've also tried to point out some questions about functionalism that are yet to receive satisfactory answers. For example, recent attention to the central notions of realization and multiple realization show how much work remains to be done in even articulating a full theory, much less evaluating its application. And these issues appear to have consequences for metaphysics, explanation, and special sciences generally, not only psychology.

8. References

- Bechtel, W. and J. Mundale. 1999. Multiple realizability revisited: Linking cognitive and neural states. *Philosophy of Science*, 66: 175-207.
- Bickle, J. 1996. New wave psychophysical reductionism and the methodological caveats. *Philosophy and Phenomenological Research*, 56, 1: 57-78.
- Bickle, J. 1998. *Psychoneural Reduction: The New Wave*. Cambridge, MA: The MIT Press.
- Bickle, J. 2006. Multiple Realizability. *The Stanford Encyclopedia of Philosophy (Fall 2006 Edition)*, Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/fall2006/entries/multiple-realizability/>.
- Block, N. 1978. Troubles with functionalism. C. W. Savage (ed.), *Minnesota Studies in the Philosophy of Science, Vol. IX* (Minneapolis, MN: University of Minnesota Press). Reprinted in Block (1980).
- Block, N. (ed.) 1980. *Readings in Philosophy of Psychology, Volume One*. Cambridge, MA: Harvard University Press.
- Block, N. 1990. Inverted Earth. In James Tomberlin (ed.) *Philosophical Perspectives 4, Action Theory and Philosophy of Mind*: 53-79.
- Chalmers, D. 1996. *The Conscious Mind: In Search of a Fundamental Theory*. New York: Oxford University Press.
- Chomsky, N. 1965. *Aspects of the Theory of Syntax*. Cambridge, MA: The MIT Press.
- Churchland, P. M. 1982. Is 'Thinker' a Natural Kind? *Dialogue* 21, 2: 223-238.
- Churchland, P. M. 1988. *Matter and Consciousness, Revised Edition*. Cambridge, MA: The MIT Press.

- Churchland, P. S. 1983. Consciousness: The transmutation of a concept. *Pacific Philosophical Quarterly*, 64: 80-93.
- Cummins, R. 1975. Functional analysis. *The Journal of Philosophy* LXXII, 20: 741-765.
- Cummins, R. 1983. *The Nature of Psychological Explanation*. Cambridge, MA: The MIT Press.
- Cummins, R. 1989. *Meaning and Mental Representation*. Cambridge, MA: The MIT Press.
- Dennett, D. 1971. Intentional systems. *The Journal of Philosophy*, 68: 87-106.
- Dennett, D. 1981. *Brainstorms*. Cambridge, MA: Bradford Books/The MIT Press.
- Dretske, F. 1995. *Naturalizing the Mind*. Cambridge, MA: The MIT Press.
- Fodor, J. 1974. Special Sciences, or the Disunity of Science as a Working Hypothesis. *Synthese* 28: 97-115.
- Fodor, J. 1986. Banish DisContent. In *Language, Mind, and Logic*, J. Butterfield (Ed.) (New York: Cambridge University Press, 1986.)
- Gillett, C. 2002. The Dimensions of Realization: A Critique of the Standard View, *Analysis* 64, 4: 316-323.
- Gillett, C. 2003. The Metaphysics of Realization, Multiple Realizability, and the Special Sciences, *Journal of Philosophy*, Nov. 2003: 591-603.
- Guttenplan, S. (ed). 1994. *A Companion to the Philosophy of Mind*. Oxford, UK: Blackwell Publishers.
- Jackson, F. 1982. Epiphenomenal Qualia. *The Philosophical Quarterly*, 32, 127: 127-136.
- Jackson, F. 1998. *From Metaphysics to Ethics: A Defense of Conceptual Analysis*. Oxford: Oxford University Press.
- Kim, J. 1993. *Supervenience and Mind*. New York: Cambridge University Press.

- Kim, J. 1998. *Mind in a Physical World: An Essay on the Mind-Body Problem and Mental Causation*. Cambridge, MA: MIT Press.
- Kim, J. 2005. *Physicalism, Or Something Near Enough*. Princeton, NJ: Princeton University Press.
- Kirk, R. 1974. Zombies v. Materialists. *Proceedings of the Aristotelian Society*, 48: 135-152.
- Kitcher, P. 1982. 1953 and All That: A Tale of Two Sciences. *The Philosophical Review* 93: 335-373.
- Levine, J. 1983. Materialism and Qualia: The Explanatory Gap. *Pacific Philosophical Quarterly*, 64: 354-361.
- Lewis, D. 1966. An Argument for the Identity Theory. *Journal of Philosophy*, 63: 17-25.
- Lewis, D. 1970. How to Define Theoretical Terms. *Journal of Philosophy* 68: 203-211. In Lewis (1983).
- Lewis, D. 1972. Psychophysical and Theoretical Identifications. *The Australasian Journal of Philosophy* 50: 249-258. Reprinted in Lewis (1999).
- Lewis, D. 1983. *Philosophical Papers, Volume I*. New York: Oxford University Press.
- Lewis, D. 1994. Lewis, David: Reduction of Mind. In Guttenplan (1994).
- Lewis, D. 1999. *Papers in Metaphysics and Epistemology*. New York: Cambridge University Press.
- Lycan, W. 1987. *Consciousness*. Cambridge, MA: The MIT Press.
- Marcel, T. and Bisiach, E. 1988. *Consciousness in Contemporary Science*. New York: Oxford University Press.
- Marr, D. 1982. *Vision: A computational investigation into the human representation and processing of visual information*. San Francisco: W.H. Freeman.

- Melnyk, A. 2003. *A Physicalist Manifesto: Thoroughly Modern Materialism*. Cambridge: Cambridge University Press.
- Miller, G. 1956. The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information. *The Psychological Review*, vol. 63: 81-97.
- Millikan, R. 1984. *Language, Thought, and Other Biological Categories*. Cambridge, MA: The MIT Press.
- Poland, J. 1994. *Physicalism: The Philosophical Foundations*. New York: Oxford University Press.
- Polger, T. 2000. Zombies Explained. In *Dennett's Philosophy: A Comprehensive Assessment*, D. Ross, A. Brook, and D. Thompson (eds). Cambridge, MA: The MIT Press
- Polger, T. 2004. *Natural Minds*, Cambridge, MA: The MIT Press.
- Polger, T. in press. Realization and the Metaphysics of Mind. *Australasian Journal of Philosophy*, 85.
- Putnam, H. 1967. The Nature of Mental States. Reprinted in H. Putnam (Ed.) *Mind, Language and Reality: Philosophical Papers, Volume 2*. (New York: Cambridge University Press, 1975).
- Putnam, H. 1975a. *Mind, Language and Reality: Philosophical Papers, Volume 2*. New York: Cambridge University Press.
- Putnam, H. 1975b. Philosophy and Our Mental Life, in H. Putnam, *Mind, Language and Reality: Philosophical Papers, Volume 2*. New York: Cambridge University Press.
- Putnam, H. 1988. *Representation and Reality*. Cambridge, MA: The MIT Press.
- Richardson, R. 1979. Functionalism and Reductionism. *Philosophy of Science* 46: 533-558.

- Searle, J. 1990. Is the Brain a Digital Computer?, *Proceedings and Addresses of the American Philosophical Association* 64: 21-37.
- Shapiro, L. 2000. Multiple Realizations. *Journal of Philosophy*, 97, 635-654.
- Shapiro, L. 2004. *The Mind Incarnate*. Cambridge, MA: The MIT Press.
- Shepard, R and J. Metzler. 1971. Mental rotation of three dimensional objects. *Science*, 171(972):701-3.
- Shoemaker, S. 1981. Some Varieties of Functionalism. *Philosophical Topics* 12, 1: 83-118.
- Shoemaker, S. 1982. The Inverted Spectrum. *Journal of Philosophy*, LXXIX, 7: 357-381.
- Shoemaker, S. 2001. Realization and Mental Causation. In Gillett, C. and B. Loewer (eds.), *Physicalism and Its Discontents*. Cambridge, UK: Cambridge University Press (2001).
- Shoemaker, S. 2003. Realization, Micro-Realization, and Coincidence. *Philosophy and Phenomenological Research*, LXVII (1): 1-23.
- Turing, A.M. 1950. Computing machinery and Intelligence. *Mind*, 59: 433-460.
- Van Gulick, R. 1980. Functionalism, Information, and Content. *Nature and System* 2: 139-162.
- Van Gulick, R. 1983. Functionalism as a Theory of Mind. *Philosophy Research Archives*: 185-204.
- Van Gulick, R. 1988. Consciousness, Intrinsic Intentionality, and Self-Understanding Machines. In Marcel and Bisiach (1988).
- Wilson, R. 2001. Two Views of Realization. *Philosophical Studies*, 104: 1-30.
- Wittgenstein, L. 1953. *Philosophical investigations*. (Anscombe, G.E.M., trans.). Oxford: Basil Blackwell.

Suggested Readings

Block, N. 1978. Troubles with functionalism. In C. W. Savage (ed.), *Minnesota Studies in the Philosophy of Science, Vol. IX* (Minneapolis, MN: University of Minnesota Press).

Reprinted in Block (1980).

Fodor, J. 1974. Special Sciences, or the Disunity of Science as a Working Hypothesis. *Synthese* 28: 97-115.

Kim, J. 1996. *Philosophy of Mind*. Boulder, CO: Westview.

Loar, B. 1981. *Mind and meaning*. New York: Cambridge University Press.

Lycan, W. 1987. *Consciousness*. Cambridge, MA: The MIT Press.

Polger, T. 2004. *Natural Minds*. Cambridge, MA: The MIT Press.

Putnam, H. 1967. The Nature of Mental States. Reprinted in H. Putnam (Ed.) *Mind, Language and Reality: Philosophical Papers, Volume 2*. (New York: Cambridge University Press, 1975).

Rey, G. 1997. *Contemporary Philosophy of Mind*. Boston: Blackwell Publishers.

Shoemaker, S. 1981. Some Varieties of Functionalism. *Philosophical Topics* 12, 1: 83-118.

Van Gulick, R. 1983. Functionalism as a Theory of Mind. *Philosophy Research Archives*: 185-204.

Glossary Terms

Autonomy

Brentano's Thesis

Epiphenomena

Etiological function

Identity theory

Intentionality

Realization

Reduction

Teleological function

Turing machine

Turing test