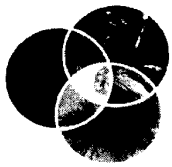
Advanced Review



# Functionalism as a philosophical theory of the cognitive sciences

Thomas W. Polger*

**Functionalism is a philosophical theory (or family of theories) concerning the nature of mental states. According to functionalism psychological/cognitive states are essentially functional states of whole systems. Functionalism characterizes psychological states essentially according to what they do, by their relations to stimulus inputs and behavioral outputs as well as their relations to other psychological and nonpsychological internal states of a system. The central constructive relation for functionalism is the so-called realization relation. Realization is a proposal for how psychological states can be real, physical, and causally efficacious while at the same time preserving the autonomy of cognitive explanations and avoiding reduction or elimination.** © 2012 John Wiley & Sons, Ltd.

How to cite this article:
*WIREs Cogn Sci* 2012, 3:337–348. doi: 10.1002/wcs.1170

## INTRODUCTION

Functionalism is a philosophical theory (or family of theories) concerning the nature of mental states. According to functionalism psychological/cognitive states are essentially functional states of whole systems.[a]

Saying that psychological states are functional states, the functionalist claims more than that psychological states have functions. Rather, functionalism is the theory that psychological states are defined and constituted by their functions. On this view, what it is to be a psychological state of a certain sort *just is* and *consists entirely* of having a certain function. Anything that has that function in a suitable system would therefore be that psychological state. If storing information for later use is the essential function of memory, then anything that has that function counts as a memory. Similarly, one might say that anything that traps or kills mice counts as a mouse trap.

The functionalist approach was first explicitly introduced by Hilary Putnam.[1,b] Functionalism characterizes psychological states according to what they do, by their relations to stimulus inputs and behavioral outputs as well as their relations to other psychological and nonpsychological internal states of a system. It is most recognizable in its computational variation, according to which psychological/cognitive states are computational states of whole systems, for example, paradigmatically, human beings, and other organisms.

## THE GENERAL IDEA OF FUNCTIONALISM

Functionalism is an 'ontological' or 'metaphysical' theory.[c] That is, functionalism is a philosophical theory about what kinds of 'things' psychological states are, about their essential natures. It is not intended to be a competitor with concrete empirical hypotheses about some or all cognitive capacities, although it may be more compatible with some than others.[d] Functionalism is an empirical meta-theory, a hypothesis about what kinds of things psychology studies. Similarly, for example, the view that there is genuine uncertainty in nature (the Heisenberg interpretation of quantum physics) is an empirical meta-theory that is intended to be compatible with rather than competing with concrete theories of quantum phenomena. As it happens, some meta-theorizing in physics is done by physicists and some by philosophers; whereas most meta-theorizing in psychology is done by philosophers, less by psychologists and cognitive scientists. But this is a mainly sociological fact about the disciplines and not any deep difference in the meta-theories of physics compared to the cognitive sciences. Not all theories in philosophy of mind

*Correspondence to: thomas.polger@uc.edu
Department of Philosophy, University of Cincinnati, Cincinnati, OH, USA

are bits of meta-psychology, but functionalism as I understand it is one such bit of meta-science.

Functionalism is intended to be a middle-ground approach, contrasted with behaviorism on one end of the spectrum and brain-specific theories on the other. Behaviorism is maybe the most familiar psychological meta-theory, according to which 'Psychology is the science of behavior' and thus the 'things' studied by psychologists are behaviors.[2] Classical behaviorists denied that psychological states were internal states of systems, preferring to construe psychological states as total conditions of systems rather than parts of them. These total states were supposed to be characterized by the stimulus inputs to the whole system, and behavioral outputs (or dispositions for output) of the whole system. Functionalism differs from behaviorism, then, by allowing that psychological states are internal states of systems. Functionalism allows that the inputs and outputs that characterize psychological states can be inputs from and outputs to and from other internal parts of the system—viz., subpersonal psychological states—in addition to stimulus inputs from the environment and overt behavioral outputs.

The functionalist approach also differs from brain-based meta-theories because it denies that the inputs and outputs from psychological states can be characterized entirely in neuroscientific terms, for example, in terms of firing rates, neurotransmitter releases or bindings, or so on. The functionalist holds that—in principle if not in fact—psychological states can be had by systems that do not have brains like those of human beings or other terrestrial organisms. Therefore, they conclude, it would be incorrect to characterize psychological states in a way that limits them to human or known neurological inputs and outputs. In principle, at least, even nonbiological systems could have psychological states, according to functionalists. The functionalist approach attempts to characterize mental states in a way that is more abstract (i.e., less specific) than brain sciences, but not so abstract (i.e., generic) as gross behavior.

Functionalism is compatible with the views that psychological systems are computing systems or information processing systems, but it is possible to endorse those views without endorsing functionalism. For example, someone might hold that human psychology is computational but also hold that computation is a purely behavioral phenomenon. Alan Turing's[3] famous 'Turing test' for machine intelligence is an example: Turing suggests that the question, 'Can machines think?' can be operationalized as the question, 'Can a finite state machine play the imitation game?' His answer strongly suggests that thinking is a computational process. But Turing's criteria are purely behavioral, and they do not require (or even much suggest) that the internal states of finite state machines should be identified with psychological states.

Similarly, the idea that psychological or cognitive systems are information processing systems does not require that any internal states of information processing systems are themselves to be identified with psychological states. For example, one could take the view that memory is an information processing function that allows a system to store information about a past event in order to produce a behavior at a later time. But that information processing view of memory does not require that the internally stored informational states be recognized as particular 'memories' or 'beliefs', that is, as psychological states. Many philosophers and cognitive scientists have been attracted to functionalism because it seems to dovetail nicely with computational cognitive science, and because it seems to have built-in resources for addressing philosophical problems about representation and mental content. But I distinguish those representationalist views from functionalism because making use of their resources does not require that we adopt the functionalist ontology.

Functionalism is more than just the claim that cognitive systems are computing systems of some sort, or that cognition is an information processing activity. First, functionalism goes beyond those views by also claiming that the internal states of the systems are in fact psychological or cognitive states, not merely mechanisms for behavior. Second, functionalism is not limited to those views because it can construe functional states in myriad ways, not only in terms of computation or information processing. In fact the most common current versions of functionalism do not overtly make use of computational systems; and some functionalists employ evolutionary notions of function rather than computational notions.

## THE VIRTUES OF FUNCTIONALISM

Functionalist theories claim to have four features that make them more plausible than other approaches to the nature of psychological states.

First, functionalism is a *realist* theory of psychological states. Functionalism identifies psychological states with functional states of whole systems, so given that we are realists about systems, for example, organisms, we can be realists about the states of those systems. Realism is a good feature for psychological theories to have because psychological states certainly appear to be real, so it matches our first-person or introspective experience, and also our common sense practices of attributing psychological

states to one another and other creatures. Perhaps more importantly, reality may be a necessary feature of psychological states if they are to figure in certain kinds of explanations and have certain kinds of effects. The contrast here is not the view that psychological states are bits of fantasy or hallucination, but rather that they are instrumental apparatus or useful fictions. For example, we may say that a truck tipped over because its center of gravity was too high, but we do not expect any entity called a 'center of gravity' to figure in the full explanation of the event, much less to actually cause the tipping. Most theorists are not realists about centers of gravity, as many of us are about beliefs and memories.

Overtly eliminativist theories that deny the reality of psychological states have never been particularly popular. But behaviorism is an antirealist theory insofar as it attributes psychological states to the whole system rather than any part of it. Behaviorists denied that there is any less-than-total state of the system that is a psychological state and that is the cause of behavior—for example, no state of thirst that causes drinking behavior, to use an example from Skinner.[4] (Functionalists, in contrast, are happy to recognize psychological states that are states of portions of the system rather than of the total system, even if they are individuated by their relations to the total system.) And many philosophers and psychologists have worried that brain-based theories have eliminativist consequences, and that they would explain ('reduce') psychological states by replacing them with brain states or eliminating them altogether. Indeed, some advocates of reductionist approaches have endorsed and advertised that consequence.[5–8] Other theorists who support the brain-based approach have maintained that like functionalism it is a realist approach.[9–11]

Second, functionalism is compatible with *physicalism*, the view that everything is broadly physical, or (for present purposes) that there are no fundamentally mental things.[12] Functionalism is also compatible with nonphysicalist accounts, for better or for worse. As Hilary Putnam says memorably, as far as functionalism is concerned the mind could be 'copper, soul, or cheese'.[13] But by being at least compatible with physicalism, functionalism denies the need for any special mental stuff—*res cogitans*, thinking stuff—in order to account for psychological states. So functionalism at least does not require any nonphysical stuff or properties. None of the major competitors to functionalism requires the introduction of nonphysical stuffs, so this virtue is not unique to functionalism. However recently there has been a resurgence of interest in theories that include nonphysical properties that are (nevertheless) somehow had by physical stuff—particularly theories of the nature of conscious psychological states such as sensory experiences.[14]

Compatibility with physicalism is a good feature for meta-theories of psychology to have because it appears to be true, and because nonphysicalist theories that are also realist face some serious obstacles. Physicalism appears to be true because no empirical theories of physical, chemical, biological, and social phenomena introduce any entities that are neither strictly physical (roughly, part of physics) nor made out of physical parts. Biologists no longer suppose that there is any *élan vital* that distinguishes organic from inorganic systems, and psychologists do not rely on the special properties of *res cogitans* in their explanations. This is good, moreover, because we have increasing reason to believe the physical world is causally closed, that there are no causal effects that come into or exit from the physical world.[15] So if there are any nonphysical things, they would have to be causally inert and therefore of limited explanatory use. Why posit such things? This is, as it has always been, the big problem for nonphysicalist theories: even if we had a reason to posit nonphysical things, we have no explanation for how they could do anything, and thus no expectation that they would be explanatorily fertile. It would not be very plausible to let the meta-theory drive the first order theory and insist that the world must not be causally closed, after all. So even contemporary property dualists—those who hold that there are extra nonphysical properties, even if there is no extra nonphysical stuff—have to admit that those properties must be epiphenomenal.[14] Most theorists regard this as reason enough to reject those property dualist theories, even if they are unsure just how a physicalist theory can succeed.[e] This leads us directly to the next virtue of functionalism.

The third attractive feature of functionalism is that it seems to show how psychological states can make a causal difference in the world. If they are real and they are broadly physical in that they are somehow or other made out of physical stuff, then psychological states will be as *causally efficacious* as any other broadly physical states, for example, chemical and biological states. This is good for at least two reasons. One is that, like realism, it seems to match our introspective experiences and common sense attributions. It certainly seems as though I walked downstairs and poured a brown liquid into my mug because I desired some coffee, because I believed that there was coffee in the carafe, and because I remembered how to get to the kitchen. That is, my psychological states of belief, desire, and memory seem to play a causal role in producing my behavior.

And I often suppose that other people's actions are caused by their beliefs, as when I surmise from the fact that my wife has brought an umbrella that she believes it might rain. Beyond these personal-level examples, many psychological and cognitive theories attribute causal powers to internal representational and information-storing states of systems. They might explain my coffee-seeking behavior, or the behavior of a rat in a maze, by citing the causal effects of a mental map that has been stored and maintained, for example.

The other, and related, reason that it is a good thing to construe psychological states as causally efficacious is that it may be a necessary condition on their appearing in true psychological explanations. If it is not true that my beliefs and desires or mental maps (or any other psychological states) actually caused me to arrive in my kitchen, then it is unclear how they could be useful in explaining how or why I arrived there. Of course talk of psychological states might still appear in some gloss of a causal explanation of my behavior, in the same way that centers of gravity can appear in glosses of causal explanations of trucks tipping over. But this is a deflationary view of psychological states and psychological explanations. On this view psychological explanations are not strictly speaking true. All else being equal it would be preferable to have a meta-theory according to which psychological states are causally efficacious and psychological explanations are true, and functionalism is one such theory. Behaviorist and overtly eliminativist theories deny the causal efficacy of psychological states because they are not realist about those states to begin with. And dualist theories must either deny the causal efficacy of the psychological or else implausibly reject the causal closure of the physical. So brain-based theories are the main competitors to functionalism that also attribute causal potency to psychological states.

Finally, functionalism explains how psychological explanations can be both true and have a certain kind of independence or *autonomy* from the explanations provided by other sciences.[1,16] Nonrealist approaches to psychological states undermine psychological explanation by denying the reality of psychological states. Approaches that deny causal efficacy to psychological states also undermine psychological explanation because they prevent psychological explanations from being true causal explanations. And nonphysicalist approaches, insofar as they must deny causal efficacy, have the same consequences. Brain-based theories do not have those problems—they are realist, physicalist, and attribute causal potency to psychological states. But brain-based theories take the connection between psychological states and brain states to be very intimate—usually identifying them with one another, in the same way that temperature states of gases are identified with mean molecular kinetic energy states of the aggregates of molecules that make up gases. Consequently, brain-based approaches deny that psychological explanations are autonomous from neuroscientific explanations, seemingly linking the success or failure of psychological explanations to the question of whether a corresponding neural explanation will be found. Likewise, explanations of gas temperature are dependent on molecular kinetic explanations, one might say. Yet many theorists have thought that psychological explanations are independent of or autonomous from neuroscience. Taking a familiar example: David Marr's[17] explanation of early vision in terms of the construction of 2D sketches via edge detection did not depend on finding retinal ganglion cells that are sensitive to zero crossings, and the model was indeed implemented in silicon computing machines that did not have any cells at all. Marr's explanation is not now widely accepted, and this is despite the fact that retinal ganglion cells do something that is not too badly characterized as detecting zero crossings.[f] This seems to show that psychological explanations neither stand nor fall by their connections to neuroscientific explanations.

A virtue of functionalism, then, is that it seems to account for how psychological states can be implemented by brains without hindering the autonomous practice of psychology. The explanation is that this can happen because psychological states are relatively abstract or general, capable of being made of 'copper, soul, or cheese'. Behaviorism also allows for an autonomous psychology, but because it denies realism to psychological states it must deny that psychological explanations are causal explanations—thirst does not cause drinking. Dualist theories might also allow for autonomous psychological explanations, but they definitely could not be causal.

Behaviorism as both a concrete psychological theory ('medthological' behaviorism) and philosophical meta-theory ('logical' behaviorism) did not enjoy much enthusiasm in the later years of the 20th century because the family of behaviorist approaches was widely thought to have been subjected to devastating critiques from both philosophers and cognitive scientists.[18,19] Eliminativism was never very popular. And dualism has been long out of favor. That leaves functionalism and brain-based theories.

In this context, the attraction of functionalism is plain. According to the functionalist meta-theory, psychological states are real, physical, causally potent, and figure in autonomous explanations. The nearest

competitor is the brain-based approach, but it denies that psychological explanations are autonomous from neuroscientific explanations. All of the other approaches are worse off because they reject at least two of the four virtues of functionalism.

## MAKING UP FUNCTIONAL STATES: REALIZATION

The trick to functionalism's allure is that it seems to explain how psychological states can have the first three virtues discussed—reality, physicalism, and causal efficacy—while also preserving the fourth, the autonomy of psychology. And the key to this metatheoretical 'home run' is the functionalist account of how psychological states can be somehow made up by brains without being identical to states of brains.

According to functionalism, psychological states are realized by but not identical to states of brains.[g] So the 'making up' relation for functionalism is *realization*. Realization for functionalism is a technical term; and recently there has been an active dispute over exactly how to characterize the realization relation, in general.[20–23] But realization for functionalism may prove to be a special case, and it is easy to understand by way of examples. The core example of realization is that of a computing machine: realization is the relation between hardware and software. Here I sit, typing an article about functionalism on my computer using a word processing program. But at least in principle if not in practice, this same program could be run on a slightly or significantly different computer—one with a different central processing chip, one with multiple processors, one built from vacuum tubes rather than solid state transistors, or one built from organic materials. The computational states of the word processing program—the stored information that the words should be displayed in a certain font, for example—cannot be identified with, say, the internal electromagnetic states of the physical device sitting on my desk right now. They cannot be identified because the same program could be run on—realized by—many different devices. And, indeed, later today this same machine might use different internal states (e.g., different memory registers) to store the same information; and it may use the same internal states to store different information, if I am running a different program. Realization differs from other 'making up' relations because it is a many-to-many relation: many different internal states of different machines can realize one and the same computational program state; and many program states could be realized by the same internal states.

The main traditional argument for functionalism is precisely that brain states appear to be in a many-to-one relation to psychological states—many different kinds of brains (or nonbrains, potentially) seem to be plausible candidates for having psychological states. According to this line of reasoning, we have empirical reason to think that psychological states are not uniquely made up (one-to-one) but are rather 'multiply realized' by different brain states in different creatures. This was Putnam's original argument for his functionalist hypothesis, and it remains the dominant reason that philosophers reject the competing brain state theory. If this 'multiple realization' reasoning is correct, then any one-to-one relation between brain states and psychological states is ruled out, including the relation of identity.[h] So the brain-to-psychology relation, whatever it is, had better be many-to-one because And functionalism is a proposal for just such a many-to-one relation, it hypothesizes that the brain-to-psychology relation is realization.

Explicitly computational versions of functionalism are presently out of favor. Much more common are versions of functionalism that replace the computational program with a psychological theory. The idea then, is that cognitive systems are those that realize psychology, and that having a psychological state is a matter of being a total system that realizes a system characterized by a psychological theory and also having an internal functional state that is a functional state of the system according to that psychological theory. Most commonly the psychological theory is pictured as a set of causal laws of psychology, and the idea is that the laws constitute psychological states.[i] In the same way, one might imagine, physical theory is the conjunction of all the laws of physics so that anything that behaved according to (a certain subset of) those laws would necessarily be—that is, realize—an electron.

In the most cutting edge forms of functionalism, the causal relations of a psychological theory are supplemented with historical and evolutionary relations ('functions') so that the causal states must also be produced in the correct ways, usually by etiological processes of natural selection, development, or learning.[24–28] The details need not concern us here. The central idea is that psychological states are functional states of the systems that have them, where those states are realized by but not identical to, for example, brain states in human beings. Computational, psychological, and teleological (i.e., etiological) ways of developing functionalism are just the most prominent proposals for how to flesh out the 'functions' that are said to be realized by physical systems, in virtue of which the physical systems have psychological states.

But all is not wine and roses. Fleshing out functionalism and functional realization has proven to

be more difficult than it originally seemed. Recall that functionalism aims to provide a middle ground between on the one hand behaviorist theories that were so abstract that they denied reality to psychological states, and on the other hand brain-based theories that were so concrete that only things with brains just like ours could qualify as having psychological states. The trouble is that this balancing act is hard to pull off. Versions of functionalism, such as computational theories, that are abstract enough to cover a variety of actual and possible psychological systems tend to attribute psychological states to things that are not usually considered good candidates—like thermostats and fuel gauges. These versions are too 'liberal' in their attribution of psychological states. But versions of functionalism, like teleological versions that require psychological systems to have a certain history of natural selection, avoid liberalism at the cost of requiring psychological systems to be almost exactly like human beings, so it is said that such theories are overly 'chauvinistic' in their attribution of psychological states. Because formulating functionalism requires deciding what sets of inputs and outputs are constitutive of psychological states, Ned Block[29] calls the problem of finding a functionalist theory that is neither overly liberal nor overly chauvinistic 'the problem of inputs and outputs.' And there is no generally agreed-upon solution to the problem.

This conundrum can also be thought of in terms of the generality or domain of psychology: Should psychology be only the science of human cognition (or of a subset of human cognition), or should it be the science of all possible cognitive systems? Worse than the mismatch with our pretheoretical expectations about the attributions of psychological states and the generality of psychological explanations, the problem of inputs and outputs reveals a tension in the claim that functionalism has all four of the virtues discussed above. Suppose the realizers of psychological states are hugely diverse, and in particular that they fall under a wide and heterogeneous range of causal laws. Then the list of possible realizers will be a disjunction, just a big list of A, or B, or C, . . . , and so on. This is particularly easy to imagine in computational versions of functionalism. In this case, it is clear why psychological kinds would not be in a one-to-one relation to the kinds of other sciences, and therefore why there would be a certain kind of autonomy of psychology as a science. But if the realizers of psychological states are so causally heterogeneous, then it is hard to see how they could be causally unified enough to figure in causal explanations. Think of all the different kinds of mouse traps. Mouse traps are so diverse that they really have nothing in common—despite the name, they do not even all trap mice. Some mouse traps trap mice, some break their necks; some poison mice, some electrocute mice; some simply divert mice. There are lots of kinds of mouse traps, but consequently there are no interesting generalizations about mouse traps out of which to form a science of mouse traps. The list of mouse traps is just a big disjunction. And so it might be with widely diverse realizers of psychological state kinds; they might turn out to be just a heterogenous grab bag of different states sharing little in common.

Suppose, instead, that the realizers of psychological states are not so diverse after all. Suppose that there are perfectly good generalizations about the realizers, just as there are perfectly good generalizations that cover the various isotopes of chemical elements. In that case, we can see how we can have a science of psychology. But it is less clear that science is autonomous from brain sciences. After all, chemical isotopes are not wildly diverse, they are variations on a core commonality, so they are plausibly in one-to-one relations to gross substance kinds rather than many-to-one relations. In that case, like temperature and mean molecular kinetic energy in a gas, it seems like the explanations might be intimately related and not independent, after all. Particular temperature states occur in various gases, but they do so in the same basic way. If psychology is like that, then psychological and brain sciences are more intimately connected than the functionalists prefer to think. There would be psychological generalizations, but they would not be autonomous.

This version of the problem of inputs and outputs is closely related to what Jaegwon Kim calls 'Descartes' Revenge'.[30] The problem with classical dualism, as mentioned above, was that it does not have an account of how psychological states can have causal efficacy. The problem of inputs and outputs shows that physicalist theories have the same kind of problem if they want to avoid identifying psychological and brain states. Either psychological and brain states are really different kinds, related many-to-one, in which case it is hard to see how psychological states can have causal powers. Or else psychological and brain states can be identified, related one-to-one, in which case causal power is assured but it's hard to see how psychological and cognitive sciences could be autonomous from brain sciences.[30]

There is a third and related problem, as well. This is that, given the causal closure of the physical, if psychological states are not identified with brain states then it is hard to see how they could fail to be epiphenomenal. It may seem that there is simply not any causal work left for them to do, if they are realized by but not identical to physical states of brains.

This 'causal exclusion' problem seems to show that if psychological states are not identical to physical states then they are either epiphenomenal or causally redundant, and those are not very attractive options.[31,32]

These problems—the problem of inputs and outputs, the disjunction problem, and the causal exclusion problem—have kept philosophers busy. While there are no generally accepted solutions, it is still fair to say that through the turn of the millennium most philosophers believed that all three problems have solutions, and that those solutions will be basically functionalist in form.[f] Given the alternatives there was little choice but to hope. Behaviorism was a failed program. Eliminativism was never very attractive. Dualism is not a live option for most contemporary theorists. And multiple realization seems to show that the brain state approaches are doomed. Functionalism seemed to be the only game in town.

## MORE TROUBLES FOR FUNCTIONALISM

The problem of inputs and outputs, the disjunction problem, and the causal exclusion problem are 'internal' problems for functionalism. They are problems, that is, that arise within the constraints of the functionalist theory, that make it seem that no version of the theory could achieve its goals while being consistent. There are also two serious 'external' objections to functionalism, objections that question the adequacy of the theory even if it can be given a consistent formulation. The first is that functionalism seems, to many, to be a poor theory of conscious mental states, such as sensations. The second is that the empirical evidence for multiple realization has increasingly come into question.

Functionalism, because it identifies psychological states in terms of what they do, works best for those kinds of psychological states and processes that are most familiar for what they do: belief, memory, perception, comprehension, and so on. But one aspect of psychological life is more salient for what it is than for what it does, namely conscious experience. Whatever it is that conscious experiences do, it seems possible that some non-conscious state could do the same job. Indeed there are countless examples in which some nonconscious state actually does the same job or nearly the same job, either normally or in some pathological condition. Probably you are reading this article using conscious visual perception. But blindsight patients show that at least some visual perception can be unconscious.[33] Frequently we experience the odors we smell, but some olfaction may occur unconsciously via the vomeronasal sense.[34] In these actual cases, the acuity and range of unconscious perception is degraded. But many theorists believe that these practical limits are incidental. There does not seem to be anything in principle that stops nonconscious visual perception from being every bit as good as normal conscious visual perception. If there were conscious and nonconscious mechanisms of visual perception that did all the same jobs and had all the same inputs and outputs—both to the world and to other mental states—then functionalism would say that those two mechanisms are instances of the same kind of psychological state. Yet it seems plain to many that conscious and unconscious mechanisms, even if they both do the same job, are different psychological states. There is a world of difference between conscious visual perception and blindsight, after all. What psychological difference could be more obvious than the difference between states that are conscious and those that are not? Numerous philosophical thought experiments and more than a few clinical cases have been used to support the idea that such nonconscious 'twins' or 'zombies' are possible in principle and maybe in practice.[14,29,35–39] If so, the reasoning goes, then functionalism is the wrong theory of conscious mental states, for it fails to make an important and essential discrimination.

These consciousness objections are extremely hard to evaluate. The examples are intuitively gripping. But the fictional examples depend on too many assumptions, not least of which the assumption that our imagination or powers of conceiving are a good guide to what is possible. The real-world examples are often messy and never illustrate the complete parity of function that is needed to make the argument go through without residual doubts. Given that no theory of consciousness seems overwhelmingly compelling, the fact that some theory might if true have some counterintuitive consequences does not make for much of an objection. Indeed, every theory of consciousness seems to have some unexpected results. This may be a case where the winner gets the spoils, where we let the best theory tell us what to say rather than demanding that our theories conform to our pretheoretical expectations.

An entirely different concern about functionalism has to do with the motivation and empirical evidence for functionalism. Recall that the reason that functionalism seems to do somewhat better compared to the brain-based theories is that the brain-based theories apparently do not accommodate the phenomenon of multiple realization, the fact that the brain–psychology relation is many-to-one. This problem is widely taken to be fatal to brain-based theories,

and it is also leveraged by the functionalist to explain how psychology can be an autonomous science even though every particular psychological state is realized in the brain.

There has always been some resistance to the claim that brain-based theories cannot accommodate multiple realization.[8,40,41] Some have come close to arguing that problematic forms of multiple realization are impossible.[30] More recently there has been a flurry of work arguing that the functionalists have overstated the empirical evidence for multiple realization and against the brain-based theories. The general tactic is to show that multiple realization is more rare than functionalists suppose, and that traditional examples fail. This is accomplished by a combination of careful attention to empirical literature, as well as by a more nuanced understanding of what multiple realization would be like in the context of scientific explanations. The full range of responses is more than we can review here, and none of the arguments is uncontroversial. But we can survey a number of concerns that have been prominent, and some that are starting to gain traction.

First, it has been pointed out that many traditional examples of multiple realization involve a mismatch of 'grain' between psychology and brain sciences.[42] We can think of psychological states in a coarse grained way, such as the *perception of red*. But we can also think of psychological states as being more fine grained, such as the *perception of Pantone Red 499*. Familiar examples of multiple realization tend to construe psychological states very coarsely, considering for example *pain* or *hunger* in general. They then formulate a hypothesis that compares the coarse-grained psychological state to some very specific and fine-grained brain state, such as the old philosophical chestnut, 'pain = c-fiber firing'. The comparison of the coarse grained psychological state makes it seem plausible that some creature could have that general psychological state type (pain, of any sort at all) but without having the precise neural state (c-fiber firing), and therefore makes brain-mind identities seem unlikely. The 'pain = c-fiber firing' example was introduced into philosophy in the 1950s, and it was always intended as a simplified stand-in for actual theories rather than a hypothesis in its own right—its empirical implausibility is widely recognized. But what is less often noticed is the way that this toy example surreptitiously makes multiple realization look more plausible than it should by offering a false hypothesis with a grain mismatch.

That actual scientific hypotheses about brain-to-mind relations would involve grain mistmatches is unlikely. One reason is that they would fail basic empirical requirements for covariation and comanipulability, therefore never being candidates for genuine explanations. But there is a more important point. Just as the toy example of pain and c-fiber firing may be misleading, so too the simplification of treating psychological and brain sciences as static and completed can mislead. We use the simplifying pretense that we are able to compare two finished theories that were independently developed. But in fact we have two sciences that interact in many ways—methodologically, explanatorily, institutionally, and so on.[k] The explanations, methods, and taxonomies of these cognitive and brain sciences constantly respond to and adjust to one another. A simple example is the more or less recent splintering of 'memory' into numerous memory phenomena, based in part on independent psychological investigation and partly on discoveries from clinical patients who suffered traumatic brain injuries.[43] For at least some if not all of these cross-disciplinary interactions, the localization and identification of neurological and psychological mechanisms, both within and across organisms and species, is a working heuristic.[44] The heuristic is not perfect of course, and the presumption can be defeated. But it leads to more one-to-one correspondence than the functionalist approach expects. Moreover, when there is not a one-to-one match, then investigators face an empirical and theoretical question about whether they have a case of multiple realization or grain mismatch, or whether the case calls for adjusting one or both taxonomic schema by merging or splitting kinds.[l]

An important factor in assessing *prima facie* examples of multiple realization is whether the range of examples is genuinely alike psychologically, and whether those examples are genuinely distinct neuroscientifically. On the one hand, some purported examples of 'same' psychological states across time or creatures, particularly after corrections for grain mismatches have been made, fail to be the same after all. For example, in many examples of compensatory neural plasticity following trauma, the psychological function appears to be performed in a new brain area but is also severely degraded. And in Sur's well-known study of rewired ferrets that 'see' with their auditory cortex, the function is degraded and the auditory cortex also reorganizes to resemble a normal visual cortex.[45,46] (see also Refs 11,47). This brings us to the second factor: the psychological states have to be the same in the compared systems, and the systems themselves have to be different. But not just any difference will yield a case of multiple realization. To use an example from Larry Shapiro[20]: two corkscrews that differ only in color do not count as different realizations of corkscrew. They

are corkscrews; and they are different. But they are not different with respect to being corkscrews; they are corkscrews in exactly the same way—their corkscrew-relevant properties are the same. To get multiple realization we need to find not just differences, but relevant differences.[11,20,47–50] Shapiro argues that once all of these factors are taken into account, the hypothesis that psychological capacities are multiply realized is less likely than the hypothesis that there are substantial constraints on the neural bases for psychological states in human and other terrestrial organisms.[11] In short, the brain-based view is more likely.

Finally, the theoretical framework in which the multiple realization argument seems to favor functionalism may need to be reconsidered. As we saw above, the contemporary functionalist approach replaces the computer programs of early functionalism with scientific psychological theories. But these theories are nevertheless program-like in that they consist of a set of laws or generalizations that are taken to be definitive of the psychological states that they describe. On this picture a psychological theory is very much like a computer program, a list of instructions to be followed that govern the behavior of a psychological system just as a program governs the behavior of a machine. Each line of the program is a true statement about the operation of the machine; each line of the theory is a true statement about the operation of the system. But it may be that psychological theories are not like that at all. For one thing, it is not clear that the cognitive sciences have any grand unifying theories, at all. Plausibly psychology (like biology, some would say) has only small, local theories or generalizations and no global theories. But more broadly, it may be that scientific explanation in general does not take the axiomatic form, does not consist in sets of laws. For example, it might be that scientific explanations do not make true claims about the systems that they cover, but instead make true claims about idealized and abstract models that more or less closely resemble real-world systems.[51] This idea, roughly put, is one part of the so-called 'semantic view' of theories. If it is right, or even if psychological theories sometimes engage in this kind of idealization or abstraction, then it may explain the appearance of a many-to-one relation between brains states and psychological states.[52] It could be that psychological state kinds are idealized or abstracted kinds, an explanatorily useful way of talking about brain states while not describing them precisely.[53,54] If so, we would expect to see an apparent diversity of brain states associated with any given psychological state kind. But this would not be because psychological states are multiple realized. Instead it would be because, out of convenience, our explanatory practices suppress by abstraction and idealization the actual variability within the psychological kinds.[m] This does not show that nonidealized psychological states can, in fact, be identified with brain states; but it leaves the prospects open.

Neither of the major objections to functionalism can be considered decisive at this time, though the empirical critiques of multiple realization can fairly be said to be gaining momentum. Nevertheless, the considerable attraction of functionalism remains, fueled both by its theoretical virtues and persistent images from science fiction that seem to suggest that creatures without brains like ours could obviously have the full range of psychological states that we have.[n]

## CONCLUSION

This article aims to provide a general introduction to the main idea behind functionalism as a philosophical position about the nature of psychological states.[o] The main theoretical virtues of functionalism were described, and compared to the virtues and vices of the most familiar contemporary competitors. The defining ontological relation for functionalism—realization—was introduced; and challenges to its understanding and employment were surveyed. Two prominent objections to functionalism were introduced, but neither were judged to be decisive.[p]

## NOTES

[a]This theory should not be confused with early 20th Century 'Chicago Functionalism' as espoused by, for example, Pierce and Angell. But the views do have certain affinities.

[b]Some scholars claim that Putnam was not the first functionalist. Some attribute that honor to Aristotle,[55] or to Sellars.[56] It may be true that Putnam was not the first functionalist, at least according to a functionalist theory of who is a functionalist. But Putnam was the first to explicitly expound and defend the functionalist view.

[c]Functionalism also has derivatives that are not primarily ontological, but rather methodological, epistemological, semantic, and so on. See Refs 9,57.

In this article I use psychology as the representative cognitive science, but readers are free to substitute 'cognitive' or cognates throughout. Similarly, I speak generally about brain sciences or neurosciences, intending those to be read inclusively.

For expository convenience I will speak mainly about psychological *states* but I use 'state' generically to cover any sort of entity that would be the target of explanation: objects, events, properties, relations, processes, phenomena, and so on. I do not think that either of the above expository conveniences changes the substance of what is said.

[d]For example, at one time there was a dispute over the compatibility of functionalism with connectionist models in psychology.

[e]A referee for this journal reminds me that the property dualist could also accept a weaker version of causal closure that is compatible with dualistic properties being causal over-determiners. That is correct. But my central point here is that we have good reasons to accept a version of causal closure that the property dualist must somehow circumnavigate.

[f]But see Ref 58 for some misgivings.

[g]It is true that some philosophers have thought that realization (and, indeed, functionalism) is compatible with mind–brain identities, for example, Ref 59. However these 'identities' are usefully distinguished from those advocated by the contemporary identity theorist, not least because the 'identities' of the former are said to be contingent, like realization.

[h]Identity is a one-to-one relation. It is the relation that everything has to itself and to nothing else.

[i]See, for example, Ref 59.

[j]Many contemporary philosophers disavow functionalism but nevertheless place their bets on some physicalist theory of the nature of minds that is compatible with multiple realization, generally called 'nonreductive physicalism'. When such theories do explicitly identify psychological kinds with functional kinds, they usually appeal to realization or to supervenience. I have suggested that realization is definitive of functionalism. Supervenience is a generic covariation relation that stands in need of explanation, typically by appeal to realization. In fact I do not know of any such theory that is not broadly functionalist, owing to the flexibility of the functionalist framework as much as the intentions of the advocates of nonreductive physicalism.

[k]Maintaining for now the pretence that there is one psychological science and one brain science.

[l]For an argument that kind-splitting is less common than one might suppose, see Ref 60.

[m]For an alternative interpretation of idealization and abstraction in psychology, see Ref 54.

[n]It must be true because data would not lie.

[o]Functionalist theories have subsequently been applied to a wide range of phenomena, among them: biological traits, ethical properties, economics, chemistry, and truth itself.

[p]This author, however, is a critic of functionalism.[9,49,61]

## ACKNOWLEDGMENTS

I am grateful to Bob Barnard, Doug Keaton, and Larry Shapiro for helpful suggestions on a draft of this article, and for ongoing discussion of functionalism and realization.

## REFERENCES

1. Putnam H. The nature of mental states. In: Putnam H, ed. *Mind, Language and Reality: Philosophical Papers*. Vol. 2. New York: Cambridge University Press; 1975.
2. Watson J. Psychology as the behaviorist views it. *Psychol Rev* 1913, 20:158–177.
3. Turing AM. Computing machinery and intelligence. *Mind* 1950, 59:433–460.
4. Skinner B. *The Science of Human Behavior*. New York: Macmillan; 1953.
5. Churchland PM. Eliminative materialism and the propositional attitudes. *J Phil* 1981, 78:67–90.
6. Churchland PM. Is 'Thinker' a natural kind? *Dialogue* 1982, 21.223–238.
7. Churchland PS. Consciousness: the transmutation of a concept. *Pacific Phil Quart* 1983, 64:80–93.
8. Bickle J. *Psychoneural Reduction: The New Wave*. Cambridge, MA: MIT Press; 1998.
9. Polger T. *Natural Minds*. Cambridge, MA: The MIT Press; 2004.
10. Smart JJC. Sensations and brain processes. *Phil Rev* 1959, LXVIII:141–156.
11. Shapiro L. *The Mind Incarnate*. Cambridge, MA: The MIT Press; 2004.
12. Montero B, Papineau D. A defence of the via negativa argument for physicalism. *Analysis* 2005, 65: 233–237.
13. Putnam H. Philosophy and our mental life. In: Putnam H, ed. *Mind, Language and Reality: Philosophical Papers*. Vol. 2. New York: Cambridge University Press; 1975.

14. Chalmers D. *The Conscious Mind: In Search of a Fundamental Theory*. New York: Oxford University Press; 1996.
15. Papineau D. The rise of physicalism. In: Gillett L, ed. *Physicalism and Its Discontents*. Cambridge: Cambridge University Press; 2001.
16. Fodor J. Special Sciences, or the Disunity of Science as a Working Hypothesis. *Synthese* 1974, 28:97–115.
17. Marr D. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. San Francisco: W.H. Freeman; 1982.
18. Putnam H. Brains and behavior. In: Butler RJ, ed. *Analytical Philosophy: Second Series*. Oxford: Blackwell; 1963.
19. Chomsky N. Review of *Verbal Behavior* by B. F. Skinner. *Language* 1959, 35:26–58.
20. Shapiro L. Multiple realizations. *J Phil* 2000, 97:635–654.
21. Gillett C. The metaphysics of realization, multiple realizability, and the special sciences. *J Phil* 2003, 591–603.
22. Polger T. Realization and the metaphysics of mind. *Austral J Phil* 2007, 85:233–259.
23. Polger T, Shapiro L. Understanding the dimensions of realization. *J Phil CV* 2008, 4:213–222.
24. Van Gulick R. Functionalism, information, and content. *Nat Syst* 1980, 2:139–162.
25. Millikan R. *Language, Thought, and Other Biological Categories*. Cambridge MA: The MIT Press; 1984.
26. Lycan W. *Consciousness*. Cambridge, MA: The MIT Press; 1987.
27. Dretske F. *Explaining Behavior*. Cambridge, MA: The MIT Press; 1988.
28. Dretske F. *Naturalizing the Mind*. Cambridge, MA: The MIT Press; 1995.
29. Block N. Troubles with functionalism. In: Savage CW, ed. *Minnesota Studies in the Philosophy of Science*, Vol. IX Minneapolis, MN: University of Minnesota Press; 1978.
30. Kim J. Multiple realization and the metaphysics of reduction. *Phil Phenomenol Res* 1992, 52:1–16.
31. Kim J. *Mind in a Physical World: An Essay on the Mind-Body Problem and Mental Causation*. Cambridge, MA: MIT Press; 1998.
32. Kim J. The myth of nonreductive materialism *Proc Address Amer Phil Assoc* 1989, 63:31–47.
33. Weiskrantz L. *Blindsight: A Case Study and Implications*. New York: Oxford University Press; 1986.
34. Keeley B. Making sense of the senses: individuating modalities in humans and other animals. *J Phil* 2002, 1:1–24.
35. Kirk R. Zombies v. materialists. *Proc Aristot Soc* 1974, 48:135–152.
36. Nagel T. What is it like to be a Bat? *Phil Rev* 1974, 4:435–450.
37. Searle J. *The Rediscovery of the Mind*. Cambridge, MA: The MIT Press; 1992.
38. Jackson F. Epiphenomenal qualia. *Phil Quart* 1982, 32:127–136.
39. Shoemaker S. The inverted spectrum. *J Phil* 1982, 7:357–381.
40. Lewis D. Review of art, mind, and religion. *J Phil* 1969, 66:23–35.
41. Kim J. Phenomenal properties, psychophysical laws, and identity theory. *Monist* 1972, 56:177–192.
42. Bechtel W, Mundale J. Multiple realization revisited: linking cognitive and neural states. *Phil Sci* 1999, 66: 175–207.
43. Tulving E, Craik FIM, eds. *The Oxford Handbook of Memory*. New York, NY: Oxford University Press; 2000.
44. Bechtel W, McCauley R. Heuristic identity theory (or back to the future): the mind-body problem against the background of research strategies in cognitive neuroscience. *The Proceedings of the 21st Annual Meeting of the Cognitive Science Society*. Mahwah, NJ: Lawrence Erlbaum Associates; 1999, 67–72.
45. Sharma J, Angelucci A, Sur M. Induction of visual orientation modules in auditory cortex. *Nature* 2000, 404: 841–847.
46. von Melchner L, Pallas S, Sur M. Visual behaviour mediated by retinal projections directed to the auditory pathway. *Nature* 2000, 404:871–876.
47. Polger T. Evaluating the evidence for multiple realization. *Synthese* 2009, 167:457–472.
48. Shapiro L. How to test for multiple realization. *Phil Sci* 2008, 75:514–525.
49. Polger T. Identity theories. *Phil Comp* 2009, 4:1–13.
50. Shapiro L, Polger T. Identity, variability, and multiple realizability. In: Gozzano S, Hill C, eds. *The Mental and the Physical: New Perspectives on Type Identity*. Cambridge: Cambridge University Press.
51. Cartwright N. *How the Laws of Physics Lie*. Oxford: Oxford University Press; 1983.
52. Klein C. Multiple realizability and the semantic view of theories. *Phil Stud*, in press.
53. Klein C. An ideal solution to disputes about multiply realized kinds. *Phil Stud* 2008, 140:161–177.
54. Haug M. Abstraction and explanatory relevance, or why do the special sciences exist? *Phil Sci* 2011, 78: 1143–1155.
55. Nussbaum MC. *Aristotle's De motu animalium*. Princeton, NJ: Princeton University Press; 1978.
56. Sellars W. Empiricism and the philosophy of mind. In: Feigl S, ed. *Minnesota Studies in the Philosophy of Science, I*. Minneapolis, MN: University of Minnesota Press; 1956.

57. Polger T. Computational functionalism. In: Calvo P, Symons J, eds. *The Routledge Companion to the Philosophy of Psychology*. London: Routledge; 2008.

58. Polger T. Neural machinery and realization. *Phil Sci* 2004, 71:997–1006.

59. Lewis D. How to define theoretical terms. *J Phil* 1970, 68:203–211.

60. Craver C. Dissociable realization and kind splitting. *Phil Sci* 2004, 71:960–971.

61. Polger T. Are sensations still brain processes? *Phil Psychology* 2011, 24:1–21.

## FURTHER READING

Bickle J. Multiple realizability. In: Zalta EN, ed. *The Stanford Encyclopedia of Philosophy (Fall 2008 Edition)*; 2006. Available at: http://plato.stanford.edu/archives/fall2008/entries/multiple-realizability/.

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

Levin J. Functionalism. In: Zalta EN, ed. *The Stanford Encyclopedia of Philosophy (Summer 2010 Edition)*; 2009. Available at: http://plato.stanford.edu/archives/sum2010/entries/functionalism/.

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

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

Stoljar D. Physicalism. In: Zalta EN, ed. *The Stanford Encyclopedia of Philosophy (Fall 2009 Edition)*; 2009. Available at: http://plato.stanford.edu/archives/fall2009/entries/physicalism/.