Functions and Qualia

Diana Mertz Hsieh (diana@dianahsieh.com)
Philosophy of Mind (Phil 5300, Hanna)
12 November 2002

In recent years, much of the hope for the basic project of materialism, that of explaining mentality in purely physical terms, has been vested in materialist varieties of functionalism (Kim 1998, 73).[1] Instead of identifying mentality with either behavior or brain states as the failed approaches of behaviorism and physicalism do, functionalism identifies mental states as "consisting of a disposition to act in certain ways and to have certain mental states, given certain sensory inputs and certain mental states" (Block 2002, 94). Thus, for the materialist functionalist, "mental states are functional states and functional states are physical states … in virtue of their causal relations" (Searle 1997, 141).

This approach to the mind seems promising, as it "is supposed to combine the best features of physicalism and behaviorism, while avoiding many of their difficulties" (Searle 1997, 139). Functionalism avoids the problems of physicalism by embracing the multiple realizability argument as its central thesis. And it avoids the problems of behaviorism by not limiting itself to external inputs and outputs, but also including internal causal relations between the mental states themselves. In these ways, functionalism appears to be the best possible continuation of the materialist inquiry into the nature of the mind at the present time.

But, we must ask, does functionalism adequately account for the central feature of mentality, namely phenomenal consciousness? The answer, in short, is no. The absent and inverted qualia arguments indicate that functionalism has not accounted for the "what its like to be" of mentality. Additionally, its account of the mind simply avoids discussing what ought to be central to any properly functional explanation, namely phenomenal consciousness.

Putnam's Machine Functionalism

In the essay "The Nature of Mental States" (originally "Psychological Predicates"), Hilary Putnam put forth the "multiple realizability argument" in claiming that mentality could be realized in a wide variety of physical systems with the requisite functional organization. This argument both undermined type physicalism and established functionalism as an alternate theory of the nature of mind (Kim 1998, 73).

Putnam's argument for multiple realizability begins with the premise that "all organisms capable of feeling pain are Probabilistic Automata," i.e. Turing machines in which "the transitions between 'states' are allowed to be various probabilities" (Putnam 2002, 76, 75). For such a Probabilistic Automaton, to be in a mental state is to be in "an appropriate kind of Functional Organization" consisting of a set of "distinct states…related to one another and to the motor outputs and sensory inputs by the transition probabilities" of the Automaton's machine table (Putnam 2002, 76, 75). In other words, an entity is in a given mental state if and only if it is in the appropriate functional state. So according to machine functionalism, to be in pain is not to have particular brain state, nor to exhibit certain behavior, nor to subjectively experience unpleasant sensations, but rather to fulfill a particular function, perhaps consisting of the detection of "tissue damage" (sensory input), some sort of "escape behavior" (motor output), and

[1] In this paper, I shall only be discussing materialist varieties of functionalism, even though functionalism itself is compatible with dualism.

a future aversion to whatever caused the pain (relation to other functional states) (Kim 1998, 77). Or, as Kim explains in technical terms, "for something to have mentality…is for it to be a physically realized Turing machine of the appropriate complexity, with its mental states … identified with the internal states of the machine table" (Kim 1998, 88). Since such functional states can be realized in any number of physical systems, so too can minds be realized in any number of physical systems. Hence, minds are multiply realizable.

As a result of the multiple realizability argument, machine functionalism denies that mental states are "physical-chemical" brain states. However, brains are not irrelevant to the mentality of humans and other conscious earthly creatures as they would be in a dualist variety of functionalism. Brain are, after all, an integral part of the physical means by which an organism instantiates the functional states of mentality. However, as Kim notes, "it is our brain's computational properties, not its biological properties, that constitute our mentality" (Kim 1998, 91).

Looking to recent philosophical history, we find that this multiple realizability argument very quickly and cleanly undermined type physicalism's arguments for the identification of mental states with brain states. After all, type physicalism requires that a particular type of physical structure—and only that physical structure—underlie any given mental state. Thus, to use the standard example, if pain is identical to the firing of C-fibers, then pain must always and only be the result of firing C-fibers and never the result of any activity of any other sort of biological or physical structure. This type physicalism, Putnam notes, involves the "ambitious hypothesis" that "parallel evolution, all over the universe, [would] *always* lead to *one and the same* physical 'correlate' of pain" (Putnam 2002, 77). Moreover, for type physicalism to be true, this principle of universal physicality would have to apply to all mental states—hearing, wondering, pondering, smelling imagining, hoping and so on—not just pain (Putnam 2002, 77). Type physicalism is thus possible according to the multiple realizability argument, but just extremely unlikely.

Putnam's machine functionalism seems to avoid such worries by adopting the multiple realizability argument as its central thesis. Consequently, the advocacy of "species-independent" functional states that are "invariantly correlated" with particular mental states seems unproblematic (Putnam 2002, 78). However, Putnam may well have merely transformed the problem faced by the type physicalist rather than eliminated it, at least according to Kim's argument in *Philosophy of Mind*. Kim notes that just as type physicalism requires the physical structures underlying any mental state to be identical across species, so machine functionalism requires the "input and output conditions" of any mental state to be "appropriately similar, if not identical" across species (Kim 1998, 95). The problem is not, as with behaviorism, that octopi must curse and swear as we often do to be considered in pain. After all, in functionalism, mental properties are abstract, not concrete (Kim 1998, 77). The problem is that of finding "an abstract enough *nonmental* description of pain behavior that is appropriate for humans and octopuses and all other pain-capable organisms and systems" (Kim 1998, 96). In other words, we must be able to find the octopus-equivalent of cursing and swearing in order for the functional output of octopus pain to the same as that of human pain, i.e. for octopi and humans to be in the same mental state at all. So machine functionalism seems to have endorsed a rather ambitious empirical hypothesis of its own.

This difficulty faced by machine functionalism, however, is not shared with another variety of functionalism, namely Armstrong's causal role functionalism.

Armstrong's Causal Role Functionalism

In the essay "The Causal Theory of the Mind," David Armstrong advances a different form of functionalism called causal role functionalism. On this theory, a mental state is "a state that is apt to be the cause of certain effects or apt to be the effects of certain causes" (Armstrong 2002, 82). The causes of mental states are "objects and events in a person's environment," while the effects of mental states are "certain patterns of behavior" (Armstrong 2002, 82). While this description may sound like behaviorism, Armstrong does not regard all of the outputs and inputs as external to the organism like a behaviorist would. Internal relations between mental states also constitute inputs and outputs such that mentality consists of "the entire network of casual relations involving all psychological states" (Kim 1998, 105; Armstrong 2002, 83).

Armstrong's argument for causal role functionalism largely rests upon an analogy between physical-functional concepts like "poison" and "brittleness" and apparently mental-functional concepts like "hunger" and "purpose" (Armstrong 2002, 82-3). To be a poison, Armstrong points out, is to have a particular sort of negative biological effect upon a living body, not to possess any particular sort of underlying chemical structure (Armstrong 2002, 82). Similarly, to say that something is "brittle" tells us about its "disposition to break and shatter when sharply struck" (Armstrong 2002, 82). Both of these concepts are functional in that they describe causes and effects rather than an underlying physical nature. Similarly, argues Armstrong, the concept "hunger" designates "a state of a person or animal that characteristically brings about food-seeking and food-consuming behavior by that person or animal" (Armstrong 2002, 82). That mental concepts are functional is "most obvious and plausible" for Armstrong in the case of the concept "purpose" (Armstrong 2002, 82). Purposes are "information-sensitive causes" that "direct behavior by using perceptions and beliefs" (Armstrong 2002, 83). As such, they are not merely behaviorist inputs and outputs through a black box of consciousness, but require explanation in terms of other mental states. All mental concepts, Armstrong argues, will involve such a "complex and interlocking set of causal factors" that form the whole psychology of the conscious organism (Armstrong 2002, 83).

Unlike in Putnam's machine functionalism, the network of causal relations for any species of creature is identified with its brain in causal role functionalism (Armstrong 2002, 82). Thus while the human, the octopi, and the android may all share the requisite functional organization for pain, their underlying physical structures may differ significantly. Pain for a human may be firing C-fibers, while pain for an octopus may be firing O-fibers, while pain for an android may be electrical impulses through pain channels. As a result of this species-specific functionalism, Armstrong avoids the problems of developing a universal functionalist psychology faced by the machine functionalist. Yet we might wonder whether Armstrong has simply leaped from the frying pan into the fire in viewing mental states as a set of disjunctive physical states, given the objections to disjunctive properties faced by type physicalism (Putnam 2002, 77; Kim 1998, 117). It appears not, for Armstrong's disjunction is unified by an underlying universal causal function, while the disjunction of type physicalism is unified by nothing except perhaps a covert and illicit appeal (according to materialism) to subjective experience.

However, Armstrong faces a different sort of problem, namely the dilemma of avoiding circularity in his account of mental states while at the same time avoiding the requirement of total psychological similarity for mental states to be the same. For Armstrong, "to be a mental state is to be an internal state serving as a causal intermediary between sensory inputs and mental states as causes, on one hand, and behaviors and other mental states as effects, on the other"

(Kim 1998, 105). In other words, mental states are defined circularly, partially in reference to other mental states. In order to avoid the circularity, Armstrong speaks of the "complex and interlocking set of causal factors," i.e. the "entire network of causal relations involving all psychological states" (Armstrong 2002, 83; Kim 1998, 83). This appeal to a creature's total psychology eliminates the circularity because, in the end, all inputs and outputs will be external. However, such a solution generates the problematic requirement that for two metal states to be the same the total psychologies of the two creatures must be the same. In other words, if we are to appeal to a creature's total psychology in explaining what a mental state is, then we must also appeal to total psychology in determining whether two mental states are the same or not. But the unpalatable result of such an appeal is that two mental states are unlikely to ever be the same, even within the same species—and perhaps not even within the same creature.

The Absent and Inverted Qualia Arguments

The most common objections to both machine and causal role functionalism concern functionalism's account—or rather apparent lack thereof—of phenomenal consciousness, particularly qualia. The "absent qualia argument" and the "inverted qualia argument" are perhaps the two most common such arguments.

In his essay "Troubles with Functionalism," Ned Block presented the absent qualia argument through a thought experiment designed to elicit doubts about whether a perfectly functional replica of a person would possess phenomenal consciousness. In the thought experiment, Block asks his readers to imagine a "homunculi-headed robot" in which either a vast number of tiny people residing in the cavity of the robot's head or the entire population of China remotely connected implement the "reasonably adequate machine table that describes you" (Block 2002, 96). By stipulation, such a Turing machine "satisfies a certain type of functional description—one that you *also* satisfy, and one that, according to functionalism, justifies attributions of mentality" (Block 2002, 97). But, Block invites us to wonder, does such a system actually have mental states, phenomenal consciousness, qualia, and the like—or not? Is there "anything it is like to be the homunculi-headed system"? (Block 2002, 97) Any doubts as to the mentality of such a system, Block argues, indicates the thought experiment is a "prima facie counter example to (machine) functionalism" (Block 2002, 97). Mental states cannot be identified with functional states if a given functional state might give rise to a mental state in some cases but not in others.

A good functionalist, however, will simply deny the validity of the intuitions pumped by Block's thought experiment. Daniel Dennett, for example, does precisely that in rejecting John Searle's Chinese Room argument, which Dennett sees as analogous to the absent qualia (or Chinese Nation) argument (Dennett 1991, 435-439). Dennett argues that the erroneous intuitions result from our "imagining too simple a case" (Dennett 1991, 438). If we properly imagined the homunculi-headed robot in all its complexity, we would see that the complexity of the "System" as whole would give rise to genuine mental states (Dennett 1991, 439). However, as Kim as points out, such a reply begs the question, for it "assumes that qualia supervene on physical/neural states, but this supervenience assumption is what is at issue" (Kim 1998, 115).

A similar objection to (particularly causal role) functionalism's account of phenomenal consciousness is found in the inverted qualia argument. Here, the thought experiment simply asks us to wonder whether two people's subjective experiences of color are, in fact, the same. In other words, could my green look like red to you and vice versa? Although Dennett characterizes inverted qualia as "a philosophical fantasy," the real possibility of qualia inversion

through pseudonormal vision belies such a dismissal. In pseudonormal vision, the photopigments in the red and green cones of a person's eyes have been reversed through the combination of two genetic abnormalities (Nida-Rümelin 2002, 100). For such a person, the subjective sensations of red and green would be reversed, even though the person would behave exactly as a person with normal color vision would. So the person with pseudonormal vision and the person with normal vision would be functionally equivalent, the same in all their inputs and outputs, but not subjectively equivalent (Nida-Rümelin 2002, 101-2; Kim 1998, 114). The obvious conclusion from this thought experiment is that functionalism offers us an inadequate account of the subjective experience of phenomenal consciousness.

Similar to the reply to the absent qualia argument, Dennett's functionalist response to the inverted qualia argument avoids the question by essentially denying qualia as subjective experience altogether. He argues that to claim qualia to be more than the total of a person's "reactive dispositions" is to cling to "the tenacious myth of the Cartesian Theater" (Dennett 1991, 398). So "the idea of holding the qualia constant while adjusting the dispositions is self-contradictory" (Dennett 1991, 398). Thus Dennett is avoiding the issue by denying the subjective nature of qualia altogether—perhaps thereby proving the anti-functionalist's point.

In sum, the absent and inverted qualia arguments seem to indicate that "by identifying sensory events with causal roles, functionalism appears to miss their qualitative aspects altogether" (Kim 1998, 114). However, arguments on both sides of the debate between the functionalists and the anti-functionalists seem to quickly degenerate into warring intuitions over difficult-to-imagine thought experiments. But perhaps we can get some leverage on the issue if we examine the nature and adequacy of causal explanations themselves.

The Limits of Causal Explanations

One possible objection to functionalism is that it offers only a partial and thus inadequate account of the mind in characterizing it purely in terms of causes and effects. After all, to fully understand some class X usually requires that we understand not just the causal powers of Xs, but also the shared properties (abstract or concrete) that give rise to those causal powers. So for example, students in a chemistry class don't simply learn the functional account of Hydrogen as the atom to which some subset of other atoms can bind; they also learn the underlying property of Hydrogen that accounts for such bindings, namely that of having a single electron.

In fact, two significant problems often emerge with attempts to offer purely causal accounts of a concept: (1) a failure to properly designate the referents of the concept and (2) an unnecessary limitation on our understanding of those referents. For example, we might offer a purely causal analysis of the concept "cake," such that the causes of cake are mixing and baking and the effects of cake are icing and eating. But such an analysis is incomplete because it doesn't tell us what cakes are or why they have such causal powers. Even a complete specification of all the causal powers of cakes would not give us these important details. As a result of this incompleteness, we cannot distinguish cakes from brownies, cobblers, and cinnamon buns, as those baked good have the same causal powers as cake. Additionally, this purely causal account doesn't even hint at some of the other important features of cake, such as that cake layers may be tiered and that cakes may stick to the pan. A laundry list of inputs and outputs does not constitute good understanding.

Strangely enough, a causal analysis of cakes is inadequate even though cakes are multiply realizable, albeit in a limited sense. Although most cakes are made of flour and sugar, some are flourless and fructose-sweetened. Some cakes use leveners while others do not. Cakes may be

made in any number of shapes, with any number of flavorings, and so on. So although we cannot make cakes out of beer cans and string, we are not straightjacketed into one particular physical realization of cake. Thus it seems that multiple realizability does not make purely causal accounts of something adequate accounts of that thing.

The puzzle here is that multiple realizability seems like an excellent criterion for functional concepts. Functional concepts ought to be ones in which the particular underlying "material cause" of something is irrelevant, in the sense that the material cause may be one of any number of things—within some range of possibilities. This understanding of functional concepts fits well with Armstrong's own analyses of the functional concepts "poison" and "brittleness" (Armstrong 2002, 82). Poison, according to Armstrong is something that "when introduced into an organism causes that organism to sicken and/or die" (Armstrong 2002, 82). But, as Armstrong notes, not any harmful substance (such as molten lead) will do, as the harmful substance must work "in a biological as opposed to a purely physical way" (Armstrong 2002, 82). Thus poison can be made of any number of physical materials, but not any physical material at all. According to Armstrong, the essence of such functional concepts is the "active" or "passive" powers of the referents of the concepts.

The problem for the purely causal account of the mind that Armstrong goes on to develop is that it is inconsistent with his earlier understanding of functional concepts. In short, the active and passive powers of something are not equivalent to the causes and effects of that thing. Why not? Because the talk of causes and effects of something ignores the central activity of that something which constitute its passive and active powers. So, for example, we cannot understand brittleness if we explain it simply in terms of being sharply struck (the cause) and transforming into shards (the effect). If we do that, we miss the all-important central activity of the brittle object actually shattering. So a causal account of something seems to be different from a functional account of something, contrary to the functionalist presumption of equivalence.

So returning to mentality, the central activity of the mind seems to be the irreducible subjective experience of phenomenal consciousness. To give a proper functional account of the mind, then, would be to discuss the mind in terms of its active and passive power for conscious experience. But, of course, this characterization of the mind would run contrary to the functionalist project of eliminating phenomenal consciousness from the discussion of mentality altogether.

The functionalist may, of course, argue that the central activity of the mind is not phenomenal consciousness but rather computation. However, this characterization would be problematic for it fails to exclude all sorts of non-mental entities, from adding machines to web servers to my fingers, from the category of minded beings. What in fact distinguishes minded beings from such computing devices seems to nothing more complicated than phenomenal consciousness. So the functionalist cannot appeal to computation as the central activity of the mind in order to give a genuinely functional account of it.

Concluding Thoughts

Functionalism's account of the mind in purely causal terms seems untenable. As the absent and inverted qualia arguments point out, functionalism fails to account for the qualia of phenomenal consciousness. And as my analysis of functional concepts indicates, functionalism's causal account of mind is not the same as a genuinely functional account of mind, which would have to include discussion of phenomenal consciousness as the central power

of the mind. Functionalism, it seems, does not offer materialism as much promise as was originally expected.

References

Armstrong, DM. 2002. "The Causal Theory of the Mind." *Philosophy of Mind: Classical and Contemporary Readings*. Edited by David Chalmers. New York: Oxford UP.

Block, Ned. 2002. "Troubles with Functionalism." *Philosophy of Mind: Classical and Contemporary Readings*. Edited by David Chalmers. New York: Oxford UP.

Dennett, Daniel. 1991. *Consciousness Explained*. Boston: Little, Brown, and Company.

Kim, Jaegwon. 1998. *Philosophy of Mind*. Boulder: Westview Press.

Nida-Rümelin, Martine. 2002. "Pseudonormal Vision: An Actual Case of Qualia Inversion." *Philosophy of Mind: Classical and Contemporary Readings*. Edited by David Chalmers. New York: Oxford UP.

Putnam, Hilary. 2002. "The Nature of Mental States." *Philosophy of Mind: Classical and Contemporary Readings*. Edited by David Chalmers. New York: Oxford UP.

Searle, John. 1997. *The Mystery of Consciousness*. New York: New York Review of Books.

Searle, John. 1992. *The Rediscovery of the Mind*. Cambridge: The MIT Press.