

Is Functionalism a Physicalist Theory of Consciousness?

Functionalism is generally taken to be a non-reductive physicalist theory about mental events. Some philosophers, most prominently Jaegwon Kim, have argued that functionalism cannot successfully claim a middle ground between dualism and reductive physicalism. I will argue that while these arguments do not successfully abolish the possibility of non-reductive physicalism, they restrain the kinds of functional models that are permissible.

Functionalism's claim of physicalism

The doctrine of Functionalism is, briefly put, that while the lower level processes of the brain are best described by physical accounts, the higher order function can only be described in terms of inputs and outputs and some sort of calculating mechanism that transforms the inputs into the outputs. Not only are such functional models said to completely capture the higher order functioning of the brain, they are said to do so in a way that an ordinary physical account could not accomplish. The functional account is consistent with the metaphysical possibility of multiple realizability. The physical mechanism by which the function is actuated is said to be irrelevant. If the wet-ware of our brains were entirely replaced by computer chips, the nature of our consciousness would be the same, so long as the same functional parameters were satisfied. Thus the true nature of consciousness (and other brain states) is not in their physical realization but in their functional properties. And while these functional properties are not identified with any one physical state, they are nevertheless said to be physical.

The analogy is often made to a computer program (Block 1997, for instance). A computer program cannot be reduced to the actual material circuitry of the computer. We could certainly imagine a computer realized by hydraulic or mechanical processes, and we would want to say that the program was one and the same program if run on the hydraulic computer. And while the computer program is undeniably physical, it is not identified with any particular physical state, but rather with the higher order functional relationships between physical states.

Mental causation

Jaegwon Kim and his supporters object to the notion that the mental can be considered causally effective. To clarify things from the start, if the mental is not causally effective, then, arguably, functionalism is as good as dead. Claiming that functional states are non-reducible entities gives them a real status. By saying they relate to each other through functional networks (of some kind or other) is to ascribe causal efficacy to them. If mental causation is impossible, then functional states are not causally efficacious. Kim (1998) argues that something that isn't causally effective cannot be said to be real. And even if you argue that real things could, in principle, be causally inefficacious, functionalism is seriously crippled as an epiphenomenalism of sorts. So if the argument against mental causations succeeds, functionalists will be forced either to deny the causal closure of the material world and say that their functional entities are in fact immaterial entities or deny that their functional entities are non-reductive. Both moves, in effect, abandon functionalism.

The argument derives from claims about what it means to be physicalist. Lynne Rudder Baker (1993) proposes that physicalism necessitates two principles that together exclude mental causation. The first is the casual closure of the physical, which Baker defines as “Every instantiation of a micro-physical property that has a cause at t has a complete micro-physical cause at t .” The second is strong supervenience, which Baker defines as “Necessarily, for any instantiation of any property, F , there is an instantiation of a micro-physical property, G , and, necessarily, anything that has G has F ”.

The problem for mental causation arises from intuitions about causal explanation. Baker offers the example that while Moby Dick being a whale was causally relevant toward Captain Ahab’s interest in him, his being a mammal was not. Likewise, if measles cause a certain birth defect and measles are a kind of red spot, we would not say that red spots had caused the birth defect. From this Baker argues that a higher level more general causal description is never valid when a lower one suffices. So this becomes not only an argument against mental causation, but against all macro causation in general.

Somebody who wants to defend macro-physical causation of any kind (including mental causation) has three lines of attack: deny casual closure of the physical (CCP), deny strong supervenience (SS), or deny that the CCP and SS negate the possibility of macro-physical causation. The discussion about these possibilities often revolves around the so-called special sciences. The thinking is that geology, biology, and maybe even economics identify real laws of nature that are not reducible to microphysical laws. Without digressing too much into the philosophy of these sciences and of laws in general, we can briefly consider whether this really what is at stake. I have some serious doubts

about whether the special sciences are really after the strict nomic relationships that microphysics discovers, and it appears that some philosophers have serious misconceptions about what sort of things are asserted by the special sciences. I don't think that arguments to the effect of "the special sciences exist, therefore non-reductive physicalism is true" need to be taken very seriously¹. Nor should opponents of non-reductive physicalism suggest that success of special science is merely apparent², for this at best conflates two notions of law. I think rather the issue is whether there are macrophysical laws. Certainly, real world scientists describe macrophysical laws, such as those governing planetary motion and those controlling plate tectonics. But the question is whether or not these are in fact hideously complex microphysical relationships, fundamentally governed by microphysical laws.

CCP, as formulated by Baker, can be denied relatively simply. We simply argue that some microphysical properties have macrophysical causes. Intuitively, this doesn't seem that difficult. Let us say that there are two magnetic substances attracted to one another. Within one of these substances there is a non-magnetic impurity. The elements of this impurity are being drawn toward the other magnetic substance. It seems impossible to account for this microphysically. The movement of the impurity occurs as a result of the movement of the magnet as a whole and because of the macrophysical magnetic fields.

Baker's formulation of SS can either be refuted along similar lines or reduced to a weaker claim of supervenience, depending on how you read it. It is at least plausible that many macrophysical properties are essentially defined in reference or relation to other

¹ *cf* Fodor, 1997

² As Baker does.

macrophysical properties, such that if those other physical properties did not exist, they would not exist. So, then whatever microphysical structure realizes those referential macrophysical properties is by itself incomplete. An example of this might be the motion of a planet, only definable in reference to the gravitational forces exerted by other heavenly bodies. Baker might answer that the microphysical structure of any other relevant bodies is part of the root microphysical reality. But this starts to look a lot like the much weaker claim of global supervenience, that two worlds which are microphysically indistinguishable are indistinguishable in all respects. Baker explicitly states that global supervenience is insufficient to rule out mental (or macrophysical) causation.

Kim does not use the same argument as Baker to reject the possibility of mental causation, nor does he explicitly reject it. He merely argues that if mental causation is possible, dualism is necessary (Kim, 1993). His argument, briefly, is that downward causation, in which higher level non-reducible properties affect the real properties of microphysical entities, is only possible if the higher properties are not causally constituted by microphysical properties (Kim, 1993). Or, in other words, upward determination excludes downward causation. The basic argument for this is that if the downwardly causing entities were microphysically determined, the microphysics would be causally effective. If the macrophysics were also causally effective, then we would have a case of causal overdetermination, which Kim (1993) claims is absurd. Though, I don't agree that causal overdetermination is necessarily absurd, its application here is pretty unreasonable. It makes little sense to say that the macrophysics also caused an effect, when we have a complete microphysical cause. Kim also claims that if mental

properties are not upwardly determined by microphysical properties (and therefore are allowed to have downward causation), then they are not physical entities.

Kim's account brings us straight into the same supervenience problems we encountered with Baker's account. Macrophysical entities are completely defined by the microphysical (global supervenience). However, this does not seem to mean that all causes must be microphysical. Tectonic and structural geology provide a wealth of examples in which parts of the earth crust causally interact. Basic mechanical laws seem to define and control these interactions. I would argue that these are examples of macrophysical effects with macrophysical causes. These macrophysical causes can downwardly cause microphysical effects. The fact that this particular silica molecule has been uplifted some tens of kilometers from deep within the crust is caused by the two tectonic plates colliding. And while these tectonic plates are constituted by microphysical entities, they cannot be identified with them. If two molecules within a tectonic plate were rearranged, then we would have one and the same tectonic plate. Only a much stronger notion of supervenience would serve Kim, and Kim needs to do a lot more work to show that mental events are not like tectonic plates.

Perhaps we could flip this around and say that the functionalists have the burden of proof. Kim offers a challenge in which functionalism needs to solve the additional question of whether there is something over and above microphysics going on in minds. I think I have shown that elsewhere in physics there are plenty of examples of macrophysical effects that have reality independent from microphysical reality. But Kim's challenge may force functionalists to show that mental entities are such macrophysical effects. The classic way functionalists have done this is, arguably, with

the claim of multiple realizability. In the next section, I will discuss Kim's response to this answer.

The basic intuition

Kim and (at least some of) his supporters do not want to deny the basic truth of multiple realizability. Rather they feel that it can be accounted for without appeal to macrophysical entities. A mental property (or any other functional property) is identified as a disjunction between the various possible physical realizers of the property. While intuitively this might seem metaphysically improbable, Kim (1992, in Fodor 1997) provides the example of the functional concept of Jade. Jade is realized by two mineral species, Jadeite and Nephrite. Jadeite is a sodium aluminum silicate, while Nephrite is a calcium magnesium hydrous silicate. Kim claims that the macrophysical properties of the two species are identical. Technically this is false; their specific gravities differ. However, for the function they serve, say in Chinese art, they are identical. Thus "being Jadeite or being Nephrite" is necessary and sufficient for being Jade.

I think I have to quibble a bit here, but be patient, the quibble is important. What exactly does "being Jadeite" mean? It doesn't mean that every molecule in the specimen is $\text{NaAlSi}_2\text{O}_6$. If there were a single molecule of some other mineral species, we would still want to call it Jade. Indeed, often the play of colors in Jade comes from different quantities of trace elements. So we need to vastly expand the disjunction to include all sorts of trace element combinations. But how much impurity can we take and still call Jadeite? Indeed, the concept of Jadeite (at least on the macrophysical level) is fuzzy on the edges, like alas so many concepts.

Suppose for a minute that we could somehow devise a non-arbitrary account of Jade. Kim could commit himself to a disjunctive definition that includes all the physically possible descriptions of Jade. Fodor (1997) argues that he would have to add all the metaphysically possible descriptions of Jade. For Fodor could certainly conceive of some ingenious way melting down, say, glass and making something with all the (relevant) macrophysical properties of Jade. He is unwilling to say whether or not this metaphysical necessity holds for the case of Jade; however he is willing to commit himself to claim that it certainly does hold for pain. For Fodor, the fundamental truth of multiple realizability is the metaphysical possibility of boundlessly many realizations. This, he argues, cannot be captured by a disjunction. I feel that Kim's difficulties stem from his insistence that the properties of Jade are fundamentally microphysical. The macrophysical properties of Jade are unproblematic.

Louise Antony and Joseph Levine (1997) point out that the disjunctive view makes many macrophysical properties unnecessarily complex. For instance, the property of having a mass of one gram has many, many possible realizers, to the point of it making very little sense to talk about mass as a disjunctive property microphysically realized. Ned Block (1997) adds the caveat that while this may be true of properties like mass, we could construe some other property, such as being dormative (all dormative things cause sleep) and the disjunctive position starts to look a little more tenable. For there is only a superficial resemblance between the set of things that cause human beings to become sleepy, not a "deep scientific" resemblance, as Block puts it. So, are mental properties like mass or like dormativeness? Both Antony and Levine and Block feel that certain mental properties are indeed like mass, while at least Block concedes that certain mental

properties are more like dormativeness. I do not have space to lay out the entire scope of these arguments. But I feel that this is the kind of discussion that Kim's challenge necessarily raises. Functionalism has additional homework in the area of showing why it is non-reductive.

Extreme Functionalisms

If, as I have argued, Kim's challenge increases functionalism's workload, then there may be certain breeds of functionalism that are unfit for the task. There is something on the order of an a priori argument for a particular model of the functional mind. The model derives from the work of Alan Turing and John von Neumann. The claim is that any sort of computation could be accomplished by a machine that reads and writes onto a tape of zeros and ones (or, in von Neumann's version, a set of two tapes, one for "memory" and one for "instructions"). So any sort of computational system is functionally equivalent to a von Neumann machine. And, inasmuch as the brain is computational in nature, it is functionally equivalent to a von Neumann machine. This approach fundamentally denies the relevance of phenomenology. My experience of a red sensation, for instance, is not computational and therefore is not functionally equivalent to a von Neumann machine. However, there are functional aspects of my red sensation, perhaps its relationship to my judgment that this book is red, or the like. Just as data is encoded into the memory of a computer, my red sensation could be encoded into some sort of belief structure, or other brain structure in which I am able to access information about a particular copy of "Consciousness Explained". This is why certain functionalists, Rey (1997) for one, are eliminativists about phenomenal consciousness. The redness of the red sensation has nothing to do with its functional properties. Dennett

(1991) is a tad more liberal. He is willing to admit that many brain processes might have nothing to do with von Neumann machines, in fact claiming that the functional equivalence of brain operations to von Neumann machines is a product of cultural evolution.

The problem with this kind of functionalism is that as it is based on an a priori investigation of the necessary conditions for things like beliefs, it is in no way actually tied to known structures of the mind. I don't know that this sort of functionalism is any better off than the a priori claim that dormative substances are dormative from their dormative nature. Unless you accept dualism, for brain states to be unique from one another, they have to be materially unique from one another. Likewise, a von Neumann machine distinguishes functionally unique mental states with a unique binary code. So according to a Kim-type reductionist, all the metaphysically possible brain states (including silicon brains, et cetera) that would realize a functionally unique von Neumann instruction could be placed in disjunction. The arguments we have advanced against this sort of move are based on appeals to macrophysical laws that could downwardly control the effects of the realizers. However, the von Neumannesque construction does not invoke any law or regularity other than the expression of certain states in terms of computer language. This seems much weaker than the necessary relationships of tectonic plates and planetary bodies or even than the macrophysical nature of Jade. It is true that all the brain states of the disjunction have a certain kind of relationship to other brain states, but there is little reason why this is any stronger than the relationship of all dormative substances to sleep.

Consciousness as special case?

If we are to preserve the non-reductive physical nature of the mental, we have to show that the regularities or laws that control the mental are indeed metaphysically real on a macrophysical level. While I do not claim to have by any means done this in this paper, I believe it is quite possible. As a conclusion, I will consider whether phenomenal consciousness is likely to play a special role in this project. The first person perspective is not inherently microphysical. Or if it is microphysical, we have no devices by which the underlying structure of our phenomenal experiences is revealed. I think we can be assured that no such device is possible. If somebody has very bad vision and sees images in a blurred fashion where someone else might see the same thing in a far more detailed fashion, we would not want to say that the person with better sight is seeing the lower-level of the experience. The first person experience (with regards to perception) is just exactly whatever you perceive in the manner you perceive it. A lower-level structure is therefore ruled out. If phenomenology is an integral part of our account of the conscious mind, we may already be committed to fundamentally macrophysical regularities in the mind. For if phenomenal consciousness is a non-reducible macrophysical description (as I have just suggested), mental properties in general may be based in macrophysical relationships.

References Cited:

- Antony, Louise; Levine, Joseph (1997) 'Reduction with Autonomy' *Philosophical Perspectives*, 11: 83-106
- Baker, Lynne Rudder (1993) 'Metaphysics and Mental Causation' in Heil and Mele (1993): 75-96
- Block, Ned (1997) 'Anti-Reductionism Slaps Back' *Philosophical Perspectives*, 11: 107-132
- Dennet, Daniel (1991) *Consciousness Explained*. Boston: Little, Brown and Co.
- Fodor, Jerry (1997) 'Special Sciences: Still Autonomous After All These Years' *Philosophical Perspectives*, 11: 149-164
- Kim, Jaegwon (1993) 'The Non-Reductivist's Troubles with Mental Causation' in Heil and Mele (1993): 189-210
- Kim, Jaegwon (1998) *Mind in a Physical World*. Massachusetts Institute of Technology
- Rey, Georges (1997) 'A Question about Consciousness' in Block, Flanagan, and Guzeldere (1997): 461-482