



Fodor on multiple realizability and nonreductive physicalism: Why the argument does not work

*(Fodor sobre realizabilidad múltiple y fisicalismo no reductivo:
por qué el argumento no funciona)*

José Luis BERMÚDEZ*¹, Arnon CAHEN²

¹ Texas A&M University

² University of Pittsburgh

ABSTRACT: This paper assesses Fodor's well-known argument from multiple realizability to nonreductive physicalism. Recent work has brought out that the empirical case for cross-species multiple realizability is weak at best and so we consider whether the argument can be rebooted using a "thin" notion of intraspecies multiple realizability, taking individual neural firing patterns to be the realizers of mental events. We agree that there are no prospects for reducing mental events to individual neural firing patterns. But there are more plausible candidates for the neural realizers of mental events out there, namely, global neural properties such as the average firing rates of neural populations, or the local field potential. The problem for Fodor's argument is that those global neural properties point towards *reductive* versions of physicalism.

KEYWORDS: multiple realizability, Fodor, reductive physicalism, nonreductive physicalism.

RESUMEN: Este artículo evalúa el conocido argumento de Fodor a partir de la realización múltiple a favor del fisicalismo no reductivo. La investigación reciente ha revelado que la evidencia empírica para la realizabilidad múltiple entre especies es débil en el mejor de los casos, por lo que consideramos si el argumento puede reiniciarse utilizando una noción "delgada" de realizabilidad múltiple intra-especie, tomando los patrones de activación neuronal individuales como los realizadores de sucesos mentales. Estamos de acuerdo en que no hay perspectivas de reducir los sucesos mentales a patrones de activación neuronal individuales. Pero hay más candidatos plausibles para los realizadores neuronales de los eventos mentales, a saber, propiedades neuronales globales como las tasas de disparo promedio de las poblaciones neuronales o el potencial de campo local. El problema para el argumento de Fodor es que esas propiedades neuronales globales apuntan hacia versiones reductivas del fisicalismo.

PALABRAS CLAVE: realizabilidad múltiple, Fodor, fisicalismo reductivo, fisicalismo no-reductivo.

* **Correspondence to:** José Luis Bermúdez. Texas A&M University. 301 YMCA Building. College Station, TX 77843-4237 – jbermudez@tamu.edu

How to cite: Bermúdez, José Luis; Cahen, Arnon. (2020). «Fodor on multiple realizability and nonreductive physicalism: Why the argument does not work»; *Theoria. An International Journal for Theory, History and Foundations of Science*, 35(1), 59-74. (https://doi.org/10.1387/theoria.20772).

Received: 23 April, 2019; Final version: 02 August, 2019.

ISSN 0495-4548 - eISSN 2171-679X / © 2020 UPV/EHU



This article is distributed under the terms of the
Creative Commons Attribution 4.0 International License

1. Fodor's Nonreductive Physicalism, Functionalism, and Multiple Realization

Among his many contributions to analytic philosophy of mind, Fodor pioneered the now widely held view known as non-reductive physicalism (Block & Fodor, 1972; Fodor, 1974, 1975). Fodor's concern in these initial papers was to maintain what he took to be *the* central tenet of physicalism, the 'generality' of physics, while retaining the *autonomy* of the special sciences (and of psychology in particular).

The 'generality of physics', according to Fodor, is "...roughly, the view that all events which fall under the laws of any science are physical events and hence fall under the laws of physics" (Fodor 1974, p. 97). With regard to psychology, a commitment to the generality of physics amounts to a commitment to a *token-identity* between psychological events and physical events (so that every event that falls under the laws of psychology is identical to a physical event that falls under some law of physics). Fodor argued that accepting token-physicalism does not entail type-physicalism; it does not entail that psychological *kinds* (or properties) are reducible to physical *kinds* (or properties). And, conversely, denying the reducibility of psychology—affirming its autonomy from physics—does not require abandoning physicalism.

The compatibility of these two claims—*generality* of physics and *autonomy* of psychology—is claimed to hold because mental kinds are *functional kinds*. That is, mental kinds (along with other, perhaps all, special science kinds) are individuated not by features of their material composition or structure, but by their functional-causal profile, where this is specified by the causal lawlike generalizations (of psychology) holding over the mental predicates that correspond to those kinds.¹ As functional kinds, psychological kinds can (metaphysically and nomologically) be realized by many different kinds of physical states. Very different types of systems, from organic brains to silicon chips might, when properly 'hooked up', have *the same* functional-causal profile that individuates a single mental kind.²

Furthermore, Fodor concludes that, as a matter of empirical fact, psychological kinds most probably *are* multiply realized. Drawing upon Putnam's (1967) 'Psychological Predicates', Block and Fodor argue that "[t]he argument against physicalism rests upon the empirical likelihood that creatures of different composition and structure, which are in no interesting sense in identical physical states, can nevertheless be in identical psychological states; hence that types of psychological states are not in correspondence with types of physical states." (Block and Fodor, 1972, p. 160) That is, recognizing that the same psychological generalizations are likely to hold of creatures vastly different in their physical composition—human beings, cats, and Martians—and perhaps indefinitely many such systems (if we consider functionally similar computational systems)—blocks any possible identification of the mental kinds appearing in such generalizations with corresponding physical

¹ "[T]he basic condition upon type identity in science" Block and Fodor tell us "is that it makes possible the articulation of the domain of laws." And, "then it looks as though substantive conditions upon type identity of psychological states will be imposed by reference to the psychological (and perhaps neurological) laws which operate upon those states and in no other way." (1972, p. 179-180)

² As Block and Fodor say: "in all probability, distinct neurological states can be functionally identical. That is, satisfaction of the criteria for type-distinctness of neurological states probably does not guarantee satisfaction of the criteria for type-distinctness of psychological states or vice versa." (1972, p. 180)

kinds. This is the multiple realizability argument against the reducibility of mental to physical kinds.

Given the centrality of this argument to Fodor's position (and other non-reductive physicalists), it has received a great deal of attention throughout the years.³ In sections 2 and 3 we address one recent and especially interesting objection to the multiple realizability of mental kinds, as advanced by Polger and Shapiro (2016). We argue that though their objections are effective against Fodor's original appeal to (inter-species) multiple realizability (section 2), he may be able to fortify his argument for non-reductive physicalism against such objections by appeal to the empirical plausibility of a more limited (intra-species) variety (section 3), based on the popular supposition that events of a single mental kind are realized by diverse neural firing patterns. However, in section 4 we argue that neural firing patterns are not the best candidates to serve as realizers of mental events. Instead, the most plausible realizers of mental events are higher-level, global, neural properties. As we bring out, this view of how mental events are realized speaks against Fodor's non-reductive physicalism and in favor of a reductive version of physicalism.

2. Polger and Shapiro on Multiple Realizability

Larry Shapiro and Tom Polger have mounted a powerful attack on the argument from multiple realizability, directed specifically against Fodor's arguments for non-reductive physicalism (e.g., Shapiro, 2000, 2008, 2010; Polger, 2002, 2004, 2009; Shapiro and Polger, 2012) culminating in their co-authored, and appropriately titled, *The Multiple Realization Book* (Polger and Shapiro, 2016).

Polger and Shapiro begin with the observation that traditional proponents of non-reductive physicalism, Fodor among them, sometimes operate with a notion of multiple realizability that illicitly considers *any* variation in would-be realizers as proof that the kind in question is multiply realizable. They agree that the world is full of variation, and that, as a result, we can expect to find a great deal of variability among different *instances* of the same natural kind. Yet, to establish that some natural kind is *multiply realizable*, variation in its instances, its would-be realizers, must be of the *right* sort. And, significantly, whether with respect to any particular natural kind the variation in its would-be realizers *is* of the right sort is an *empirical* question that cannot be resolved *a priori*.

To illustrate their main point, their favored example is the corkscrew.⁴ Corkscrews, they note, come in a great variety of different shapes, sizes, and colors, and are made of various different materials, from plastics, to wood, steel and titanium, all of which satisfy the function 'corked-bottle in → uncorked-bottle out'. Nonetheless, they argue, it would be a mistake *merely on those grounds* to suppose that the kind 'corkscrew' is multiply realizable. Variation in color and material composition among different corkscrews is *irrelevant* to whether the kind 'corkscrew' is multiply realized, as such features are irrelevant to their satisfying the

³ There is a vast literature on the multiple realizability argument. See Bickle (2019) for a review and comprehensive bibliography.

⁴ As they note, being a corkscrew isn't a *pure* functional kind, as being a member of this kind presumably requires satisfying certain material and structural constraints—at the very least involving a screw of sufficient rigidity.

corkscrew-function. Nonetheless, they argue that the functional kind ‘corkscrew’ is multiply realizable by *different kinds* of mechanisms—there are single-lever, ‘waiter’, corkscrews, there are double-lever corkscrews, and likely many other screw-involving mechanisms that when in appropriate contact with a corked bottle produce an uncorked bottle in its stead. What makes these not merely different instances of the corkscrew kind, but different realizers of the same functional kind is that each requires a different mechanical explanation of how the function that individuates the kind ‘corkscrew’ is brought about.⁵

More generally, Polger and Shapiro propose their ‘Official Recipe’ for multiple realization (2016, p. 67). For two entities, A and B, to be different realizers (rather than mere instances) of the same realized kind, and, hence, for the latter to count as multiply realized, the following criteria must be satisfied:

1. As and Bs are of the same kind in model or taxonomic system S1 (the realized kind).
2. As and Bs are of different kinds in model or taxonomic system S2 (the realizing kinds).
3. The factors that lead the As and Bs to be differently classified by S2 must be among those that lead them to be commonly classified by S1.
4. The relevant S2-variation between As and Bs must be distinct from the S1 intra-kind variation between As and Bs, rather than simply mapping onto the S1 kind variation.

Conditions (3) and (4) are the most crucial elements here. Condition (3) captures the requirement that the differences between As and Bs must be *relevant* to their both being classified as the same realized kind, while condition (4) requires that the differences between As and Bs do not merely map onto individual differences within the same functional kind, but that they contribute to their sameness with respect to the functional kind; so that As and Bs are, as they put it, ‘differently the same’. Thus, while it is clear that the length of the lever in a single-lever corkscrew is relevant to its carrying out its corkscrew function—so that different corkscrews that differ in the length of their lever would satisfy condition (3)—nonetheless, variations in the length of the lever also bring about differences in the corkscrew’s functional description—for example, how well or easily the corkscrew function is carried out. As a result, the two corkscrews would not satisfy condition (4), because they would be differently different (even though these differences are within the same broad functional kind), not differently the same.

They sum up their position:

“We have argued that multiple realization is not just variation, it is a distinctive kind of variation. This is why the mere fact that the world is Heraclitean—that all is flux—does not by itself

⁵ As Polger and Shapiro put it: “Because these two devices make use of different mechanical principles—one levers, one rack and pinions, and so forth—we conclude that they do the corkscrew job in different ways ... In contrast, if we consider two waiter’s corkscrews that differ in the material of which they are made ... or in their color ..., the imagined science of mechanical artifacts tells us that these differences are not relevant differences (Shapiro, 2000). The reason is that the same explanation of how the device does its corkscrew job applies to all of them, regardless of material composition and color.” (Polger and Shapiro, 2016, p. 65)

show that mental states are multiply realized. Multiple realization requires a special pattern of variation: Relevantly the same function performed in relevantly different ways, where the differences contribute to the sameness in function and not just to the differences in function.” (Polger and Shapiro 2016, p. 77)

With this in mind, they argue that we have little reason to think that specifically *mental* kinds are multiply realizable. Consider Fodor’s argument for the multiple realizability of mental kinds. Fodor claims that a great variety of species, both actual (human beings, cats, etc.) and imagined (e.g., Martians), can realize the same functional mental kinds. However, as Polger and Shapiro argue, whether some kind or other is multiply realizable depends on whether we have plausible empirical reasons to think that the variations in would-be realizers are of the right kind. It depends on whether ‘relevantly the same function is performed in relevantly different ways.’

And, in fact, as many have argued (Polger and Shapiro, most recently, but see also Couch, 2004), the assumption that psychological generalizations cut across different species lacks empirical grounding. Different species are not merely different with respect to their physical composition and structure; they are also *functionally* different. And, if so, we have good reasons to think that it is not the case that across different species ‘*relevantly the same functions* are performed in relevantly different ways’; thus, violating condition (1) above. Furthermore, to the extent that *similar* functions are performed by different physical systems, it is likely that the differences among these physical systems merely account for the *differences* in function, rather than their similarity. In this way, these differences would not count as different realizers of the same function, as they would violate point (4) of Polger and Shapiro’s criteria for multiple realizations above.

As one example, discussed by Couch (2004) and later by Polger and Shapiro (2016), consider the case of the camera eye, different instances of which occur in human beings and octopuses (among other species, of course). On the face of it, the camera eye appears to be multiply realized by two physically different systems that have evolved independently but are the product of convergent evolutionary forces (where the same traits are ‘selected for’ by different evolutionary pressures).⁶ As Couch notes, though both eyes involve a single lens and a retina, they differ in the kinds of pigments in their photoreceptors. Furthermore, “The retina in the human eye contains an array of rods and cones, and light is focused onto the receptors by bending the lens to change its shape. The retina in the octopus eye contains rhabdomeres, and focusing light involves moving the lens backwards and forwards within the shell. The parts of the structures are different and operate in different ways.” (Couch, 2004, p. 202) Importantly, these differences in the implementing physical mechanisms *count against* the claim that the camera eye is multiply realized, because they bring with them stark *functional* differences. For example, having a single pigment means that octopuses are color blind, and they respond to different kinds of stimuli with different reaction times, etc.

The example of camera eyes in different species is not a direct example of a *mental* kind, but given that the eye is a critical component of the physical realizer of a creature’s *perceptual* states, it is certainly relevant to the question of whether the *same* kinds of per-

⁶ Evolutionary convergence was one of Block and Fodor’s (1972) central empirical considerations in favor of the multiple realizability of mental kinds.

ceptual states are multiply realized in different species. Though at a high level of abstraction it is clear that both octopuses and human beings have perceptual states, the differences among the realizers of these perceptual states account for differences among the kinds of perceptual states realized.⁷ More generally, the fact that different species significantly differ with respect to their sensory systems (as well as their output, behavioral, systems) should already call to question the functionalist supposition that the same range of psychological generalizations will hold across these species.⁸

Another example presented by Polger and Shapiro with some detail concerns, a favorite of functionalists, the physical realizers of *pain* across species. The observation that different species of animal exhibit similar avoidance behaviors in response to noxious stimuli prompts us, along with Putnam, Block, Fodor and others, to suppose that these different species all realize the same mental kind *pain*. However, for pain to be multiply realized by these different species, the similar functional kind must be accompanied by relevant differences among its physical realizers. Yet, upon inspection, they argue, we find a great deal of relevant physical *commonality* among different creatures to which we are prone to attribute pain—physical commonality that accounts for the mental similarity. Focusing on octopuses once more, to the extent that we are willing to attribute to them pain, we find that they share similar physical mechanisms to those underlying human pains. Like human beings, octopuses too (like many others in the animal kingdom) have nociceptors, and mechanoreceptors in particular, that account for their reactions to mechanical damage. Yet, because they lack thermal nociceptors, they do not react as human beings do to cold. Thus, octopus-pain *is* different from human pain—the latter but not the former is an outcome of extreme cold—but it is also similar. And, crucially, the similarities and differences among the realized mental kinds are accounted for by the similarities and differences among their physical realizers (thus, once more, violating either condition (1) or (4) of their official recipe). Contrary to initial appearances, the empirical evidence suggests that pain is not multiply realizable between octopuses and human beings (and many other creatures).⁹

⁷ So, if we elect to individuate perceptual states coarsely, taking into account only very coarse-grained psychological generalizations, we will find that such psychological generalizations appealing to such perceptual-functional kinds do in fact hold of a variety of species. But, then, the differences among the physical realizers of those coarse-grained perceptual kinds will merely account for individual differences within the same functional kind, and in this way violate condition (4) of Polger and Shapiro's recipe for multiple realization. On the other hand, if we individuate perceptual states finely, taking into account very fine-grained psychological generalizations, we will find that such psychological generalizations do *not* hold across different species. The physical differences among realizers account for their realizing different perceptual-functional kinds, and in this way violate condition (1) of the recipe.

⁸ “[E]stablishing multiple realization requires showing that the same psychological state has diverse realizations. But we can always disagree with the functional taxonomy, and claim there are psychological differences at another level of description. So what needs to be established is that the psychological states are type identical, but it seems it will only be shown that the states have superficial resemblances.” (Couch, 2004, p. 203)

⁹ Another interesting example involving the octopus, which Polger and Shapiro discuss (2016, pp. 113–114), is memory. Both human beings and octopuses have memory, yet in the case of the former a hippocampus is involved whereas the octopus lacks a hippocampus. On the face of it the kind *memory* appears to be multiply realized. Yet, quoting from Hochner et al. (2006), Polger and Shapiro suggest that the same features of the hippocampus that realize memory in human beings are also present in the

Finally, Bechtel and Mundale (1999) argue that reflection on the actual conduct and history of cognitive science reveals that such research relies on there being both functional *and* physiological commonalities across species. Without such commonalities, animal models would be of little use in the study of human cognition. Yet, such models have historically been, and still are, immensely informative. As just a couple of examples, we can think of Mishkin, Ungerleider and Macko's (1983) research on the two visual systems, which was originally conducted on macaques, and Felleman and van Essen's (1991) identification of 32 different visual processing areas, once more, in the macaque.¹⁰ The significance of this research is its applicability to human vision, and such applicability is *premised* on the inter-species similarity between the underlying brain structures realizing similar visual processes. Such similarities are in fact found. And, the historical success of such research is good reason to take the premise to be warranted.¹¹ Thus, though the physical structures that realize, say, object identification in the macaque brain and the physical structures realizing the same perceptual process in human beings differ, it turns out that those differences aren't *relevant* differences. Object identification is not multiply realized because, as Polger and Shapiro would put it, the differences among physical realizers are not differences that contribute to the sameness of the realized function, thus violating condition (3) of their official recipe.

Taking these different examples together, we find that Fodor's claims about the likelihood of inter-species multiple realization lack empirical support. In many cases, what was initially thought of as a single, but multiply realized, mental kind turns out upon closer inspection to be different mental kinds. In such cases, it turns out that the generalizations of psychology do not hold across species, as physical differences among species are accompanied by relevant functional differences that suggest that what was initially considered a single mental kind ought to be split into different mental kinds (this is what is often termed 'kind-splitting', discussed at some length by Polger and Shapiro). As we have seen, both octopuses and human beings perceive and feel pain. Yet the physiological differences between the two species—differences in the kinds of pigments in their photoreceptors, or difference in the range of nociceptors they have—realize different kinds of perceptions and different kinds of pains; thus, violating condition (1) of Polger and Shapiro's official recipe. (Alternatively, these same physiological differences merely account for individual differ-

octopus' MSF-VL system and are responsible for realizing memory in the octopus. Though there are anatomical and physiological differences between the two systems, the systems are similar with respect to the features that are relevant to each system's realizing the same mental/cognitive kind, memory. As a result, the variability between human beings and octopuses does not give us reason to suppose that memory is multiply realized by the two kinds of systems.

¹⁰ The significant point, as they say, is that "...while van Essen's ultimate interest is in human visual processing, this work has been carried out on the macaque. The clear assumption is that the neural organization in the macaque will provide a defeasible guide to the human brain" (Bechtel and Mundale, 1999, p. 183).

¹¹ As they say: "One might think, at first glance, that the ability to make comparisons across species actually depends upon multiple realizability. In fact, it is the very similarity (or more precisely, homology) of brain structures which permits us to generalize across certain species. So, in this latter respect, in the context of neuroscientific research, they are not multiply realized" (Bechtel and Mundale, 1999, pp. 177-178).

ences within the same *coarse-grained* mental kind, in which case condition (4) is violated.) In other cases, when we have good reason to suppose that similar psychological generalizations *do* hold across different species, we find physiological *commonalities* accounting for that fact (thus violating condition (3)). Indeed, comparative neuroanatomical cognitive research requires, and its success in turn supports, the existence of such physiological similarities in functionally similar species.

There are obvious limits to what can be concluded from examples such as these. But, at a minimum, we can safely say that the argument for multiple realizability across species is far from secure. The best argument for multiple realizability would be actual examples of functional kinds that are multiply realized across different species and it is not clear that there are any examples satisfying the four criteria that Polger and Shapiro propose. Since Fodor rightly emphasized that multiple realizability is an empirical claim, this leaves his argument for nonreductive physicalism somewhat of a hostage to fortune.

But perhaps there are examples of multiple realizability holding within species, rather than across species. This would give examples of what we will term *thin multiple realizability*, as opposed to the *thick multiple realizability* that would be yielded if psychological kinds were to turn out to be multiply realizable across species. We will be exploring this possibility in the remainder of this paper, considering whether it can be used to reboot Fodor's argument for nonreductive physicalism in the face of the multiple problems that there seem to be for the original argument based on thick, cross-species multiple realizability.

3. *Using thin multiple realizability to reboot Fodor's argument against nonreductive physicalism*

As we observed in the previous section, there is little compelling evidence that any psychological capacities are multiply realized across species, and what evidence there is projects a much more complex picture than Fodor's argument from multiple realizability for nonreductive physicalism suggests. This section considers a more mundane type of multiple realizability, holding within a species rather than across species, and shows how it can be used to reboot Fodor's argument.

We begin by pointing out what we are *not* talking about. Some proponents of multiple realizability have proposed neural plasticity as a candidate source of multiple realizability. In particular, they have pointed to the phenomenon of cortical plasticity, which occurs when one cortical area adapts to perform a function or functions normally carried out by a different cortical area that has been damaged or is dysfunctional for some other reason.¹² Polger and Shapiro (2016) argue convincingly that even the most plausible candidates (such as the interesting example of ferrets "rewired" so that visual information from their right visual field is processed by the auditory cortex) fail to satisfy the four criteria in their Official Recipe, for reasons similar to those that count against candidates for cross-species

¹² See for example Barrett (2013) and Richardson (2009) for philosophical appeals to cortical plasticity as evidence of multiple realizability.

multiple realization.¹³ As with the cross-species examples, the jury is still out, of course, but to the extent that cases of neural plasticity are to supply effective support for multiple realization, they must meet the four criteria identified above, and none that we know of so far seem capable of doing so.

The type of thin multiple realizability that we will be discussing is much simpler altogether. To get the flavor of the proposal consider the following passage from Peter Menzies, describing experiments carried out by Richard Andersen and colleagues at the California Institute of Technology on how intentions to reach are encoded in the monkey motor cortex.

The neural signals that encode the monkey's intentions to reach for certain targets were recorded as averages of the firing rates (spikes per second) of individual neurons. But clearly the same aggregate firing rate in a group of neurons is consistent with a lot of variation in the behavior of individual neurons. For example, very different temporal sequences of neural firings can give rise to the same firing rate. So an intention to reach for a certain target can be realized in many different ways at the level of individual neurons. (Menzies, 2013 p. 62)

According to Menzies, therefore, a given intention to reach a specific target is multiply realized (over time) by different token patterns of neural firings.

Clearly, this way of thinking about multiple realizability is very different from that discussed, and criticized, in the previous section, as well as from cortical plasticity. And at first glance it seems immune to the general line of objection raised by Polger and Shapiro against these more complex candidates for multiple realizability, because it is assumed *ex hypothesi* that there are no significant functional differences across the different neural realizers. Indeed, in the experimental set-up described the truth of the assumption is guaranteed, because the monkey is trained to make exactly the same movement across the different trials, and what is multiply realized is the intention to make that movement.

In fact, it seems that his type of thin multiple realizability satisfies the four criteria in Polger and Shapiro's Official Recipe. Consider a situation in which a single intention to move I_7 is multiply realized within the same monkey by two different neural firing patterns, say, NFP_{26} and NFP_{47} . Here is how each of the criteria are satisfied:

- (i) NFP_{26} and NFP_{47} are both correctly classified as instances of the same intention to move, I_7 .
- (ii) NFP_{26} and NFP_{47} are both correctly classified as different neural firing patterns.
- (iii) The factors that lead NFP_{26} and NFP_{47} to be differently classified *qua* neural firing patterns are among those that lead them both to count as instances of the same intention to move. This holds because they differ in the respective firing profile of their constituent neurons, but it is their firing profiles that secure their common causal profile.
- (iv) The variation between NFP_{26} and NFP_{47} *qua* neural firing patterns is distinct from any variation that there might be between them *qua* intentions to move. This holds trivially because there are no, or only negligible differences between them *qua* intentions to move.

¹³ See Polger and Shapiro (2016, Ch. 5).

On the face of it, therefore, there seems to be at least one example of thin multiple realizability that satisfies Polger and Shapiro's Official Recipe. And if this example works, then so too must countless many others.

The question that we will tackle in the remainder of this paper is whether appealing to this type of thin multiple realizability can reboot Fodor's argument from multiple realizability to nonreductive physicalism, given the lack of substantive evidence for the stronger forms of thick multiple realizability that he himself proposed and discussed.

It will be helpful for the following to have a more detailed perspective on what Fodor thinks that nonreductive physicalism involves, and how exactly it is supported by multiple realizability. Fodor is working with a standard, Nagelian (1961), account of reduction, which requires the existence of 'bridge laws' relating predicates of the special science—here, psychology—and physical predicates. On this view, where S is some special science, say psychology, and $S_1x \rightarrow S_2x$ is a law within S , "[a] necessary and sufficient condition of the reduction of (1) [$S_1x \rightarrow S_2x$] to a law of physics is that the formulae (2) and (3) be laws, and a necessary and sufficient condition of the reduction of S to physics is that all its laws be so reducible.

$$(2a) S_1x \Leftrightarrow P_1x$$

$$(2b) S_2x \Leftrightarrow P_2x$$

$$(3) P_1x \rightarrow P_2x.$$

' P_1 ' and ' P_2 ' are supposed to be predicates of physics, and (3) is supposed to be a physical law" (Fodor 1974, p. 98).

Fodor takes the multiple realizability argument to show that there can be no such bridge laws.¹⁴ That is, given that a host of different physical systems can satisfy the same natural kind predicates of psychology—i.e., the multiple realizability considerations above—we can, at best, expect to find 'bridge laws' that relate a mental predicate to a wildly heterogeneous *disjunction of physical predicates* (of the form $Sx \Leftrightarrow P_1x$ or P_2x or P_3x or $P_4x...$). He observes: "...this is tantamount to allowing that at least some 'bridge laws' may, in fact, not turn out to be laws, since ... a necessary condition on a universal generalization being lawlike is that the predicates which constitute its antecedent and consequent should pick out natural kinds." (1974, p. 108) Yet, the wildly heterogeneous disjunction of physical predicates to which the candidate 'bridge' law refers (i.e., P_1x or P_2x or P_3x or $P_4x...$) does *not* pick a natural kind, because disjunctive predicates are not projectible (Fodor, 1974, 1997; Kim 1992).

It is straightforward to run a version of this argument using the example of thin multiple realizability above. Let's say that it is a law of psychology that an intention to act in a

¹⁴ "What I have been doubting is that there are neurological natural kinds co-extensive with psychological natural kinds. What seems increasingly clear is that, even if there is such a co-extension, it cannot be lawlike. For, it seems increasingly likely that there are nomologically possible systems other than organisms (namely, automata) which satisfy natural kind predicates in psychology, and which satisfy no neurological predicates at all. Now, as Putnam has emphasized, if there are any such systems, then there are probably vast numbers, since equivalent automata can be made out of practically anything. If this observation is correct, then there can be no serious hope that the class of automata whose psychology is effectively identical to that of some organism can be described by physical natural kind predicates...." (Fodor, 1974, p. 105)

certain way will (if the appropriate background conditions are satisfied) lead to the relevant action. We can write this as $I_7 \rightarrow A$. Suppose as before that I_7 is multiply realized by neural firing patterns NFP_{26} and NFP_{47} and that we are interested in a potential reduction of psychology to neuroscience (which is surely an essential step on any reduction of psychology to physics). Then, on the Nagelian model, a successful reduction requires a *bona fide* law of the form $I_7 \Leftrightarrow (NFP_{26} \vee NFP_{47})$. But there can be no such bona fide law because $(NFP_{26} \vee NFP_{47})$ is not a natural kind.

4. Assessing the rebooted argument from thin multiple realizability

We are prepared to grant that the rebooted argument from thin multiple realizability succeeds in showing that there can be no reduction of intentions to move to neural firing patterns—and hence that there are no prospects for reducing psychological laws involving intentions to move to neuroscientific laws involving neural firing patterns. However, we reject the claim that this shows that there can be no reduction of psychology to neuroscience. The basic problem with this line of argument is that it is mistaken as to the candidates it proposes for the neural realizers of intentions to move—and, by extension, for where in general it thinks that the neural realizers of psychological events are likely to be found. There is no reason that the realizers of psychological events will be highly specific neural events, such as actual neural firing patterns. Instead, we propose, there are far more plausible candidates for the role of neural realizers. And when we switch attention to more plausible candidates, we will see that there actually are good grounds for thinking that there exist precisely the type of laws that the rebooted argument claims do not exist.

Here is a preliminary observation. As mentioned in section 1, it is a basic tenet of Fodor's nonreductive physicalism that mental events be token-identical to their neural realizers. However, there is a simple reason for thinking that mental events cannot be token-identical to actual neural firing patterns. This is because some of the counterfactuals that are true of a given mental event fail to hold of the corresponding neural firing pattern.

As before, let I_7 be a particular intention to move; let NFP_{26} be the associated neural firing pattern in the monkey motor cortex; and let A be the relevant reaching movement. Then we can expect the following two counterfactuals to hold:

$$(4a) I_7 \Box \rightarrow A$$

$$(4b) NFP_{26} \Box \rightarrow A$$

In the language of possible worlds, (4a) says that in all the nearby worlds in which the intention I_7 occurs, so too does the action A , while (4b) says the same thing of neural firing pattern NFP_{26} . Consider, however, the following two counterfactuals:

$$(5a) \sim I_7 \Box \rightarrow \sim A$$

$$(5b) \sim NFP_{26} \Box \rightarrow \sim A$$

On the face of it, (5a) holds, but not (5b). (5a) holds because in all the nearby worlds in which the intention to move is absent, the monkey fails to perform action A . In contrast, (5b) fails to hold, because the nearby worlds in which neural firing pattern NFP_{26} does not occur are worlds in which some other neural firing pattern with the same causal profile occurs (say, NFP_{47}), bringing about action A . To put it another way, for the monkey to move

despite not intending to move, there would have to be some fairly fundamental changes either in the antecedent conditions or in the laws of nature—whereas a world in which the monkey moves in the absence of neural firing pattern NFP_{26} would differ from this world only in the firing behavior of a relatively small number of neurons.¹⁵

That neural firing patterns stand in different counterfactual relations to actions than do intentions to move (and mental events more generally) is not news. Some philosophers have drawn drastic conclusions from it. E. J. Lowe, for example, has tried to use those different counterfactual relations as an in-principle argument against any version of token-identity.¹⁶ Only slightly less ambitiously, Peter Menzies, independently and in collaboration with Christian List, has used pretty much the very example that we are discussing in a series of papers to argue against the causal closure of the physical (Menzies, 2013, 2015; List and Menzies, 2009; Menzies and List, 2010). In our view, Lowe, Menzies, and List are all over-stating the case.¹⁷ At most, all we can conclude is that specific neural firing patterns are not the right candidates to serve as the realizers of given mental events. But that should simply motivate us to look elsewhere for the appropriate realizers, not to abandon either token-identity or the causal closure of the physical.

Still, having said that, we do have a clear constraint upon any such candidate realizer, namely, that it should stand in the right kind of counterfactual relations, so that it satisfies the equivalents of both (4b) and (5b). To that end, we will argue in the following that the most plausible neural realizers for conscious mental events are not specific firing patterns, but rather higher-level, or global, properties of populations of neurons that are themselves *instantiated in* (but not necessarily realized by) specific firing patterns. Those higher-level properties stand in relations of counterfactual dependence to actions that typically mirror those of conscious mental events.

In Bermúdez and Cahen (2018) we proposed a hypothesis about the high-level properties that realize mental events. That hypothesis drew on important work on the neural correlates of the BOLD (Blood Oxygen Level Dependent) signal, which is the quantity that is directly measured in functional magnetic resonance imaging (fMRI).¹⁸ The basic physiological premise of functional neuroimaging is that cognitive activity increases cellular activity, which in turn increases blood oxygenation, so that tracking changes in levels of blood oxygen as a cognitive task is performed is a good guide to the areas of the brain engaged in performing that task. But, neuroscientists have asked, what exactly is the cellular activity that brings about increases in blood oxygenation?

The most obvious candidate, proposed by Geraint Rees, Karl Friston, and Christoph Koch among others, is that the blood oxygen level in a given neural region increases in proportion to the average firing rate of the neurons in that neural region (Rees, Friston, and Koch, 2000). Nikos Logothetis and collaborators came up with a different suggestion,

¹⁵ We are adopting a broadly Lewisian account of similarity across possible worlds, although we do not follow Lewis in adopting his strong centering requirement, which would make counterfactuals of the form (4a) and (4b) trivially true whenever the antecedent and the consequent are both true at the actual world.

¹⁶ See, for example, Lowe (2008, pp. 103-107) and Paprzycka (2014) for critical discussion.

¹⁷ We have discussed Menzies arguments in Bermúdez and Cahen (2015, 2018).

¹⁸ For a less condensed exposition of the material in this and the following paragraph see section 3.6 of Bermúdez (2019).

based on justly celebrated experiments that involved recording single cells in the monkey primary visual cortex within an fMRI scanner (Logothetis 2001; Logothetis et al, 2001).¹⁹ The experimenters were able to calibrate the strength of the BOLD signal, as measured by fMRI, with more fine-grained measures of neural activity measured through single-cell recordings. The results seemed to point to an alternative to the average firing rate hypothesis. This is the *local field potential*, which is an electrophysiological signal believed to measure the sum of *inputs* to the neurons in a given area—as opposed to the neural firing rates, which are the *outputs* from the neurons.

Both the local field potential and average firing rates are examples of global properties of neural populations that have been studied by neuroscientists. We suggest that the realizers of mental events are such *global neural properties* (GNP), as opposed to specific neural firing rates.²⁰ There is no need for the moment to identify a specific type of GNP as a candidate realizer. All we want to propose is that we should be looking for the realizers of mental events at the level of GNPs, rather than at the level of specific neural firing patterns. The important point is that GNPs stand in a many-one relation to specific neural firing patterns.²¹

Suppose, then, that the neural realizers for mental events are GNPs, as we have suggested, and consider the overarching aim of functional neuroimaging. One of the basic goals of functional neuroimaging is to identify meaningful correlations between the performance of different cognitive tasks, on the one hand, and the BOLD signal, on the other.²² Moreover, the basic aim of experimental design in imaging studies is to screen out correlations that are not law-like. Hence the emphasis on replicability and consistency across studies. Of course, the enterprise is still at a relatively early stage of development, and there are all sorts of methodological issues to tackle.²³ But still, we claim, the project of functional neuroimaging makes scientific sense to the extent that it looks for law-like correlations between different types of cognitive function (as manifested in experimental tasks) and cellular activity in the brain, as measured in the BOLD signal.

If this way of thinking about the enterprise of functional neuroimaging is correct, then we can expect a completed neuroscience to contain laws connecting different types of cognitive activity to global properties of neural populations. Let's hypothesize that neuroscience has moved on a long way, so that functional neuroimagers are able to identify not just changes in the BOLD signal, but also the underlying global neural properties. This is all that we would need to derive the required conclusions about counterfactual dependence.

¹⁹ These experiments were very complex from an engineering point of view, because the standard (metal) tools for single-cell recording cannot be used in the magnetic field of a scanner.

²⁰ In the remainder of this section we will speak in terms of mental events being realized by properties. Strictly speaking, the neural realizers are events with the relevant participating properties.

²¹ It may be more appropriate to think of the relation as instantiation, rather than realization. For one reason, as we will shortly see, the counterfactual dependence between GNPs and actions maps onto that between intentions and actions, which is not true of specific neural firing patterns. See Bermúdez and Cahen (2018, n. 16).

²² Of course, this is not the only goal. Functional neuroimagers are also greatly concerned, for example, to plot patterns of functional and effective connectivity within the brain. But arguably it is the existence of meaningful correlations between mental activity and the BOLD signal that underwrites, and gives point to, the search for connectivity.

²³ For more details see Bermúdez (2019) Ch. 9, as well as sections 3.5, 3.6, and 17.1.

Let's assume, to stick with our example, that the motor intention I_7 is operationalized through a particular task. That task might, in fact, be performing action A. After repeated, well-designed experiments in the scanner, a team of neuroscientists proposes as a law that some particular GNP, say GNP_{123} , reliably causes action A (when the background conditions are appropriately configured). GNP_{123} might be a given average firing rate for some neurons in the motor cortex, or it might be a specific value for the local field potential for that population of neurons—or some as yet unknown global neural property. Then, given that laws support counterfactuals and that performing action A is the experimental operationalization of motor intention I_7 , it will follow that the following two pairs of counterfactuals both hold:

$$(6a) I_7 \square \rightarrow A$$

$$(6b) GNP_{123} \square \rightarrow A$$

and

$$(7a) \sim I_7 \square \rightarrow \sim A$$

$$(7b) \sim GNP_{123} \square \rightarrow \sim A$$

These are exactly the counterfactuals required for GNP_{123} to count as a realizer of I_7 .

However, and here is the rub, the very reasons that make GNP_{123} a plausible candidate for the neural realizer of motor intention I_7 count towards reductive physicalism, not a Fodorean non-reductive physicalism.²⁴ This is because all the considerations that we have been reviewing point towards the existence of precisely the kind of bridge laws that yield a standard, Nagelian reduction. If there is a lawlike correlation between GNP_{123} and action A, and performing action A is the experimental operationalization of motor intention I_7 , then there seem to be no good reasons to deny the existence of a bridge law linking I_7 and GNP_{123} . But once we have those bridge laws, then we are well on the way to reductive physicalism.

5. Conclusion

For Fodor, the route to non-reductive physicalism relies on the multiple realizability of mental kinds. It is multiple realizability that guarantees that no genuine bridge laws relating mental kinds to physical kinds can be established, so that the former cannot be reduced to the latter. However, as Fodor claims, and Polger and Shapiro (2016) emphasize, the multiple realizability of the mental is an *empirical* hypothesis. As such, its plausibility depends on what evidence we might have for the *actual* multiple realization of the mental. Yet, when considering the actual evidence, things don't look good for non-reductive physicalism.

Genuine cases of cross-species (or, thick) multiple realizability, which were the focus of Fodor's original arguments, are hard to come by. Closer inspection of initially appealing

²⁴ In his (2012), Shapiro provides different motivations for thinking that the realizers of mental events may be higher-order (or, 'aggregate') neural properties, which would call into question the thesis of distinctness at the heart of non-reductive physicalism.

examples, such as the presence of pains or a capacity for memory in octopuses, reveals their appeal to be illusory.

Nonetheless, Fodor could still support non-reductive physicalism if he could establish the empirical plausibility of thin, intra-species, multiple realizability of mental kinds. Indeed, many have thought that mental kinds are multiply realized by different neural firing patterns in different individuals within a species and even in different time slices of a single individual. However, we have argued that upon closer inspection neural firing patterns are not plausible candidate realizers of mental kinds. Instead, we argued that the neural realizers of mental kinds are more plausibly higher-level, global, properties of neural assemblies. Yet, crucially, because such properties stand in the same relevant counterfactual dependence relations as do the mental kinds they realize, we have good reason to think that these mental kinds *are reducible* to those higher-level properties.

In sum, contra Fodor, empirical considerations lead us to deny the (inter-species and intra-species) multiple realizability of mental kinds, to deny non-reductive physicalism and the autonomy of psychology, and to affirm, instead, the possible reducibility of psychology to neuroscience.²⁵

REFERENCES

- Barrett, D. (2013). Multiple realizability, identity theory, and the gradual reorganization principle. *British Journal for the Philosophy of Science*, 64(2), 325-346.
- Bechtel, W. & Mundale, J. (1999). Multiple realizability revisited: Linking cognitive and neural states, *Philosophy of Science*, 66, 175-207.
- Bermúdez, J. L. (2019). *Cognitive science: An introduction to the science of the mind* (3rd Ed.). Cambridge: Cambridge University Press.
- Bermúdez, J. L. & Cahen, A. (2015). Mental causation and exclusion: Why the difference-making account of causation is no help. *Humana.Mente: Journal of Philosophical Studies*, 29, 47-68.
- Bermúdez, J. L. & Cahen, A. (2018). Mental causation and counterfactuals: a new argument for the type-identity thesis. In Mihretu G. (Ed.) *Consciousness and the ontology of properties*. New York: Routledge.
- Bickle, J. (2019). Multiple realizability. *The Stanford Encyclopedia of Philosophy (Spring 2019 Edition)*, Ed-ward N. Zalta (Ed.), URL = <https://plato.stanford.edu/archives/spr2019/entries/multiple-realizability/>
- Block, N. & Fodor, J. (1972). What psychological states are not. *Philosophical Review*, 81, 159-81.
- Couch, M. (2004). Discussion: A defense of Bechtel and Mundale. *Philosophy of Science*, 71(2), 198-204.
- Felleman, D. J., & van Essen, D. C. (1991). A distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex*, 1, 1-47.
- Fodor, J. (1974). Special sciences, or the disunity of science as a working hypothesis. *Synthese*, 28, 97-115.
- Fodor, J. (1975). *The language of thought*. New York: Thomas Crowell and Co.
- Fodor, J. (1997). Special sciences: Still autonomous after all these years. *Philosophical Perspectives*, 11, 149-63.
- Hochner, B., Shomrat, T. & Fiorito, G. (2006). The octopus: A model for a comparative analysis of the evolution of learning and memory mechanisms. *Biological Bulletin*, 210, 308-17.
- Kim, J. (1992). Multiple realization and the metaphysics of reduction. *Philosophy and Phenomenological Research*, 52, 1-26.

²⁵ We thank an anonymous referee for helpful comments on this paper. We are also grateful to the Israel Science Foundation for a grant (425/15) that supported Arnon Cahen's work on this paper.

- List, C. & Menzies, P. (2009). Nonreductive physicalism and the limits of the exclusion principle. *Journal of Philosophy*, 106, 475-502.
- Logothetis, N. K. (2001). The underpinnings of the BOLD functional magnetic resonance imaging signal. *Journal of Neuroscience*, 23, 3963-71.
- Logothetis, N. K., Pauls, J., Augath, M., Trinath, T., & Oeltermann, A. (2001). Neurophysiological investigation of the fMRI signal. *Nature*, 412, 150-7.
- Lowe, E. J. (2008). *Personal agency: The metaphysics of mind and action*. Oxford: Oxford University Press.
- Menzies, P. (2013). Mental causation in a physical world. In S. C. Gibb, E. J. Lowe, & R. Ingthorsson (Eds.). *Mental causation and ontology* (pp. 58-87). Oxford: Oxford University Press.
- Menzies, P. (2015). The causal closure argument is no threat to non-reductive physicalism. *Humana.Mente: Journal of Philosophical Studies*, 29, 21-46.
- Menzies, P. & List, C. (2010). The causal autonomy of the special sciences. In C. Macdonald & G. Macdonald (Eds.). *Emergence in mind* (pp. 108-128). Oxford: Oxford University Press.
- Mishkin, M., Ungerleider, L. G., & Macko, K. A. (1983). Object vision and spatial vision: Two cortical pathways. *Trends in Neuroscience* 6, 415-417.
- Nagel, E. (1961). *The Structure of Science*. New York: Harcourt, Brace & World.
- Paprzycka, K. (2014). Lowe's argument against the psychoneural token identity thesis. *Pacific Philosophical Quarterly*, 95(3), 372-396.
- Polger, T. (2002). Putnam's intuition. *Philosophical Studies* 109(2), 143-70.
- Polger, T. (2004). Neural machinery and realization. *Philosophy of Science* 71(5), 997-1006.
- Polger, T. (2009). Evaluating the evidence for multiple realization. *Synthese* 167(3), 457-72.
- Polger, T. & Shapiro, L. (2016). *The multiple realization book*. Oxford: Oxford University Press.
- Putnam, H. (1967). Psychological predicates. In W. H. Capitan & D. D. Merrill (Eds.). *Art, mind, and religion*. Pittsburgh, PN: University of Pittsburgh Press.
- Rees, G., Friston, K., & Koch, C. (2000). A direct quantitative relationship between the functional properties of human and macaque V5. *Nature Neuroscience*, 3, 716-23.
- Richardson, R. (2009). Multiple realization and methodological pluralism. *Synthese*, 167(3), 473-492.
- Shapiro, L. (2000). Multiple realizations. *Journal of Philosophy*, 97(12), 635-54.
- Shapiro, L. (2008). How to test for multiple realization. *Philosophy of Science*, 75(5), 514-25.
- Shapiro, L. (2010). Lessons from causal exclusion. *Philosophy and Phenomenological Research*, 81, 594-604.
- Shapiro, L. (2012). Mental manipulations and the problem of causal exclusion. *Australasian Journal of Philosophy*, 90, 507-524.
- Shapiro, L. and Polger, T. (2012). Identity, variability, and multiple realizability. In S. Gozzano & C. Hill (Eds.). *The mental and the physical: New perspectives on type identity* (pp. 284-287). Cambridge: Cambridge University Press.

JOSÉ LUIS BERMÚDEZ. Born in Bogotá, Colombia and educated at Cambridge University, José Luis Bermúdez is Professor of Philosophy and Samuel Rhea Gammon Professor of Liberal Arts at Texas A&M University, where he previously served as Dean of Liberal Arts. His books include *The Paradox of Self-Consciousness* (1998), *Thinking without Words* (2003), *Decision Theory and Rationality* (2009), *Understanding "I"* (2017) and *The Bodily Self* (2018).

ADDRESS: Texas A&M University. 301 YMCA Building. College Station, TX 77843-4237.
E-mail: jbermudez@tamu.edu

ARNON CAHEN is a visiting scholar at the Center for Philosophy of Science at the University of Pittsburgh, where he was also a visiting fellow. He received his PhD from Washington University in St. Louis, and held post-doc positions at Ben-Gurion University, Haifa University and the Open University in Israel. His research concerns the role of nonconceptual content in perception, thought, and action, as well as the metaphysics of mental causation and fundamental ontology.

ADDRESS: University of Pittsburgh, 1117 Cathedral of Learning. Pittsburgh, PA 15260.
E-mail: cahena@gmail.com