The Puzzling Resilience of Multiple Realization

Thomas W. Polger, University of Cincinnati
thomas.polger@uc.edu

Lawrence A. Shapiro, University of Wisconsin
lshapiro@wisc.edu

Penultimate draft of a paper forthcoming in
Minds & Machines

The Puzzling Resilience of Multiple Realization

Abstract: According to the multiple realization argument, mental states or processes can be realized in diverse and heterogeneous physical systems; and that fact implies that mental state or process kinds cannot be identified with particular kinds of physical states or processes. More specifically, mental processes cannot be identified with brain processes. Moreover, the argument provides a general model for the autonomy of the special sciences. The multiple realization argument is widely influential, but over the last thirty years it has also faced serious objections. Despite those objections, most philosophers regard the fact of multiple realization and the cogency of the multiple realization argument as plainly correct. Why is that? What is it about the multiple realization argument that makes it so resilient? One reason is that the multiple realization argument is deeply intertwined with a view that minds are, in some sense, computational. But we argue that the sense in which minds are computational does not support the conclusion that they are *ipso facto* multiply realized. We argue that the sense in which brains compute does not imply that brains implement multiply realizable computational processes, and it does not provide a general model for the autonomy of the special sciences.

Acknowledgements

We are grateful to two anonymous referees for comments that substantially improved this article.

1. The Multiple Realization Argument

In the beginning, Hilary Putnam created the multiple realization argument.

According to Putnam, mental states or processes can be realized or implemented by multiple and heterogeneous "physical-chemical" states or processes in different kinds of systems, just as computer programs and computational states or processes can be realized or implemented by multiple and heterogeneous "physical-chemical" states or processes in different kinds of systems. A particular word processor, web browser, or other app can be run on diverse operating systems, diverse hardware platforms, and in principle on machines made from any materials whatsoever. Putnam inferred that the same holds for mental processes.

It follows, given some plausible assumptions, that it is a mistake to identify mental states with "physical-chemical" states of brains, for identity is necessarily a one-to-one relation whereas realization is not. Similarly mistaken would be identifying the computational states of the machine on which we are now working with the specific electro-chemical processes going on in this physical device, for different electro-chemical processes in other devices can implement the same computational processes. So too, the relation between mind and brain is plainly not one-to-one because mental states can be realized in systems with heterogenous brains — such as mammals and mollusks — and perhaps by creatures without brains at all. Mental processes cannot be identified with brain processes because mental processes are *multiply realized* — or at least *multiply realizable*.

The multiple realization argument is, in the first case, an argument against the theory that psychological states or processes can be identified with brain states or processes, which Putnam calls the "brain state theory" and which is more widely known as the mind-brain type-identity theory. Putnam takes the multiple realization argument, along with other arguments regarding the failings of behaviorism, to favor his own "functionalist" hypothesis. According to his original formulation, psychological states should be identified with functional states of the whole system, where these "functional states" are states of probabilistic automata that are realized or implemented by human brains — but could be realized by different kinds of brains in different kinds of creatures, by physical structures that are not brains, and even by immaterial souls as far as that goes.

Subsequently, Putnam himself and his former student, Jerry Fodor (Fodor 1974), generalized the multiple realization argument. It's not just mental processes that should be identified with functional processes. All the processes of the "special" or non-fundamental sciences should be understood in this way, as being realized or implemented by the things and processes of more general and more fundamental sciences, all the way down to the bottom, viz., the most general and basic fundamental physics. If so, the multiple realization argument against the identity theory is simply a special case of the general argument against "reduction" of one science or theory to another (e.g., Kitcher 1984). This multiple realization argument yields a general model for explaining why we have a plurality of sciences. The plurality follows because the "special" sciences explain processes that are multiply realized by heterogeneous processes that are the subject matter of more general or basic sciences. Multiple realization leaves us with no choice but to have a plurality of sciences: "there are special sciences not because of the nature of our epistemic relation to the world, but because of the way the world is put together: not all natural kinds (not all the classes of things and events about which there are important, counterfactual supporting generalizations to make) are, or correspond to, physical natural kinds" (Fodor 1974: 113).

Philosophy, as a discipline, is famously short on widespread agreement. But if there is one exception, it is the consensus that multiple realization is both ubiquitous and important. By the early-mid 1970s the mind-brain identity theory had about as many supporters as the flat-

Earth hypothesis and the once plump hope for reducing special sciences to some single more basic science had shriveled. This consensus endures to this day.

Consensus is not unanimity, of course. There have always been pockets of resistance to the multiple realization argument. Jaegwon Kim (1972) and David Lewis (1969) were early skeptics of the multiple realization argument, suggesting that mind-brain identification is possible because suitably delimited "species-specific" psychological kinds are not multiply realized, after all. Berent Enç (1983), likewise, suggested that psychological predicates, properly understood, are to be "indexed" to human psychology. More generally, Enç argued that the interplay in our developing psychological and neurophysiological sciences virtually guarantees that their predicates will end up co-extensive—a line of reasoning revived by Bechtel and Mundale (1999). By the late 1980s, Kim had distilled his concerns into a pair of problems for "nonreductive physicalism" that are now quite prominent: a dilemma about multiple realization and causal unity, and the causal exclusion problem (see the papers collected in Kim 1993). In the 1990s, the assumptions about explanatory and theoretical reduction that drive the multiple realization argument against reduction came under increasing pressure (e.g., Bickle 1998, Sober 1999; see also Richardson 1979, 2009). And crossing into the current millennium, empirical concerns about the evidence for multiple realization became prominent (e.g., Bechtel and Mundale 1999, Shapiro 2000, 2004, 2008; Polger 2002, 2004, 2009a, 2009b; Craver 2004; Hemmo and Shenker 2015; Polger & Shapiro 2016; Thagard 2022; Maimon and Hemmo 2022; Cao 2022; for thorough reviews, see Bickle 2010 and 2020)

Despite fifty years of objections to the multiple realization argument, most philosophers continue to believe that multiple realization decisively rebuts the mind-brain identity theory and secures the "autonomy" and plurality of non-fundamental sciences. Fodor boasted:

> The conventional wisdom in philosophy of mind is that “the conventional wisdom in philosophy of mind [is] that psychological states are ‘multiply realized’... [and that this] fact refutes psychophysical reductionism once and for all. […]” Despite the consensus, however, I am strongly inclined to think that psychological states are multiply realized and that this fact refutes psychophysical reductionism once and for all. As e.e. cummings says somewhere: “nobody looses [sic] all of the time.” (Fodor 1997: 149)

Fodor’s confidence in the truth of multiple realization receives confirmation in virtually every introduction to the philosophy of mind in use today.[1] In short, the general consensus remains fixed: multiple realization is pervasive and – for that reason – reductionist projects are doomed.

Nobody should be surprised that even fifty years of objections have failed to change minds in philosophy. But shouldn’t fifty years of objections at least take the shine off the idea that the multiple realization argument is obviously correct? Why do so many philosophers embrace the multiple realization thesis?[2]

---

[1] Such claims are difficult to adjudicate, of course. But it is telling that even critics of multiple realization tend to comment on its widespread acceptance. Ned Block (1997) writes, “For nearly thirty years, there has been a consensus (at least in English-speaking countries) that reductionism is a mistake and that there are autonomous special sciences. This consensus has been based on an argument from multiple realizability” (1997: 107). Jaegwon Kim, in his *Philosophy of Mind* textbook writes, “by the mid-1970s most philosophers had abandoned reductionist physicalism, not only as a view about psychology but as a doctrine about the special sciences… All this stemmed from a single idea: the multiple realizability of mental properties” (2011: 129, italics removed.) And John Bickle, in his *Stanford Encyclopedia of Philosophy* entry on Multiple Realizability, remarks on the “the still-too-prominent view in the philosophy of mind that multiple realization spells doom for reductive materialism and mind-brain identity theory once and for all” (2020). It is telling that of the 79 essays in the second edition of David Chalmers’ *Philosophy of Mind* anthology (2021), fewer than a half dozen seriously consider views according to which multiple realization is false, and they were all originally published over 40 years ago.

[2] A referee suggested sociological reasons for philosophers’ ready acceptance of multiple realization – a tendency to jump on the bandwagon, especially when its drivers are charismatic figures like Putnam and Fodor. We do not wish to dismiss this possibility, but we choose to focus on philosophical justifications for the continuing popularity of the thesis. Why is the idea of multiple realization so easy to sell? Surely, if pressed, advocates of multiple realization would sooner appeal to philosophical justifications than to sociological ones, even if (as the referee suggests) the sociological reasons might be an important tacit cause. Our question is whether such philosophical justifications are cogent, apart from whether they are the “real” motivators.

We hypothesize that the appeal of psychological multiple realization is tightly intertwined with Putnam's original idea that cognition is in one way or another a computational process, and derives from the purported certainty that computational states or processes are themselves multiply realizable.[3] Before going further, we wish to respond to those who might wonder whether it is computationalism specifically, rather than functionalism more generally, that underlies the commitment so many philosophers make to multiple realization.[4]

Note that functionalism comes in a variety of different forms, from the austere machine functionalism of Putnam, to the more expansive computationalism of Fodor, to the teleological conceptions of functionalism we find in the work of Millikan and Dretske (cf. Polger 2004, Levin 2018). In our view, computational functionalism provides the most plausible justification for multiple realization for the simple reason that the computational states that computational functionalists identify with mental states constitute the most abstract kinds of states of any functionalist theory, and this is the reason cited by, e.g., Block and Fodor (1972). Compare, for instance, functional descriptions of the mental state *pain* in both computational and teleological terms. On the computational account, the description of pain amounts to nothing more than a pairing of inputs and outputs, described as generally as possible with respect to the physical details of the organism experiencing pain. The teleological account, by contrast, typically appeals to selection histories, where these histories, if filled out, would include facts about the organism in interaction with its environment. When filled in completely, the teleological story

---

[3] We don't deny that there may be other possible sources; we only propose that this is an important one. Some candidates for accounts of multiple realization that do not overly appeal to considerations of computational abstraction, e.g., the subset view (Wilson 1999, 2011) or the dimensioned view (Aizawa and Gillett 2009a, 2009b, 2011). But notice that both of those views are broadly reductive, and therefore do not purport to provide an account of the "autonomy" of the special sciences in the manner that Fodor requires.

[4] Though we cannot pursue the question here, we have doubts about whether motivations for functionalism are ever fully independent of considerations of multiple realizability.

would say a great deal about how a capacity for pain evolved in a given lineage. The story would include details that would make references to the organism's physical structure seemingly unavoidable. Those sorts of considerations suggest that computational functionalism, more than other forms of functionalism, offers the best hope for multiple realization because the computational states with which it identifies the functions of mental states are least committed, have least to say, about the physical nature of their realizers. If we raise doubts about whether computational functionalism implies multiple realization then we will have succeeded in calling into question multiple realization's "best hope."

We're now considering the idea that if cognition is a computational process and such processes are multiply realizable, then it's obvious that cognition is multiply realizable. How sure is this line of reasoning? Among the many philosophers who endorse the connections between cognition, computation, and multiple realization is Gualtiero Piccinini (2015, 2020). He builds his theory of physical computing around the assumption that brains literally compute and defines computing in a way that is intended to imply multiple realization. And even in philosophically distant contexts, the idea that cognition is computing is taken for granted. Kwame Anthony Appiah, discussing degrees of rational belief and the importance of idealization, assumes that cognition is computational. He writes:

> A computational theory of mental states treats beliefs as representations; and it says that some mental processes are computations with these and other representations. …What makes the account computational is that, along with probabilities and desirabilities, these representations also have *computational structures*: properties whose functional significance is that the outcome of certain mental processes—the computations—is the production of a new state whose

> computational structure is a function of the computational structures of the preexisting "input" states." (2017: 78)

The idea that thinking is a form of computation is so prevalent that it typically gets lost in the background until brought to our attention. Appiah, himself, seems to accept the implication that some sort of functionalism must be correct because mental states are computational. But does the mere fact that we can make decisions that approximate ideal rationality imply that thinking is essentially computational? And, supposing so, does this in turn imply that mental states are multiply realizable?

These are not rhetorical questions. We wish to challenge the assumption that if cognition is computing then it is also multiply realizable. There is much to be said about computation, cognition, and multiple realization. Here we will discuss three problems. First, the analogy between computing and cognition may not support the conclusion that mental states are literally computational states. Second, if mental states are computational states, that does not render them obviously multiply realizable. Third, even if mental states were computational states in a way that makes them multiply realizable, the argument still does not generalize to other special sciences so as to provide an account of the plurality of sciences. The modest conclusion is that the multiple realization argument cannot be taken for granted. In closing, we revisit a different way of thinking of the plurality of sciences, also found in Putnam and Fodor, that does not depend on mental processes being computational or on the ubiquity of multiple realization.

2. The Analogy of Computing and Cognition

The analogy between thinking and calculating was not new; but it took on a new form after Alan Turing's 1950 paper, "Computing Machinery and Intelligence," published in *Mind*, which helped introduce to philosophers the idea of a finite state automaton. In a series of papers in the 1950s-1960s, Hilary Putnam considered and developed the idea that cognition is a computational process. Whereas Turing proposed a test to sidestep the ontological question of the nature of mind, Putnam proposes an answer.

Putnam's innovation is to upgrade the analogy between thinking and computing into a theory. He generalizes Turing's definition of automata to include those that have state transitions that are probabilistic — any probability lower than 100%. Putnam then proposes the hypothesis that mental states are, simply and literally, states of that ontological kind: "I shall, in short, argue that pain is not a brain state, in the sense of a physical-chemical state of the brain (or even the whole nervous system), but another *kind* of state entirely. I propose the hypothesis that pain, or the state of being in pain, is a functional state of a whole organism" (1967 in 1975: 433). This hypothesis is Putnam's so-called functionalism or computational functionalism — the theory that mental states are to be identified with functional states of the appropriate sort, in this case, with computational states of probabilistic automata.

Putnam's computational functionalism dovetailed nicely with the prevailing theories in the cognitive and psychological sciences in the 1970s and 1980s, according to which cognition is explained in terms of information-processing. Fodor supported this proposal, on the philosophical side, with an argument for the Computational Theory of Mind, according to which cognition requires language-like symbols — to wit, the Language of Thought — in order to exhibit its formal and logical features. Fodor defended the Computational Theory of Mind as the "only game in town" (1975) for explaining those important features of cognition.

This is not the place to rehash debates over the correctness of the Computational Theory of Mind or the Language of Thought. We concur with Thomas Polger and Lawrence Shapiro

(2016), who have argued that Fodor's "only game in town" argument is no longer plausible, if ever it was.[5] At the very least, nobody should think commitment to the Computational Theory of Mind is obvious. So we regard as dubious any defense of multiple realization that emerges from the purported necessity of Fodorian Computational Theory of Mind.

But before we consider the features of computation and computers, let's take a step back. The line of reasoning from computation to multiple realization relies on the premise that cognition is *literally* a computing process, so that it actually and *literally* has the features of computing processes. It won't be enough to claim merely that cognition is similar to computing, or that cognition can be computationally modeled. From the fact that thinking is in *some* ways like computing, it does not follow that minds and brains have all the features of computers. After all, even digestion is "like" computing in some sense: it can be described as a function that takes as input *undigested-food* and produces as output *digested-food*. Surely, however, this is a poor reason for regarding stomachs as computers. Are brains any better candidates?

One reason to doubt that brains should be understood as literally implementing computational processes is that talk of computation in the brain sciences often does not specifically invoke the notions of computation that philosophers have in mind. Sometimes talk of computation is merely a placeholder: "In many quarters, especially neuroscientific ones, the term 'computation' is used, more or less, for whatever internal processes explain cognition" (Piccinini and Scarantino 2010: 244). But even when the term is used a bit more selectively, we should still be cautious. For example, Rosa Cao argues:

> The notion of computation is ubiquitous in discussions of the aims and advances of cognitive science and neuroscience. This is partly a practical matter, since

---

[5] The "only game in town" idea was assailed as early as Dennett (1983) and Kosslyn and Hatfield (1984); and then more vigorously with the advent of connectionist alternatives (Hatfield 1991).

> computational models are used in the analysis of everything from behavioral performance in cognitive tasks to massive brain imaging data sets to fine- scaled neural activities. But beyond the practical utility of computational tools, the computational approach – *computationalism* – represents a commitment to a theoretically and philosophically motivated conception of the nature of the brain as a computational system, and the nature of cognition as a computational process. Computation is often closely associated with semantic notions, and talk of "representation" is everywhere in neuroscience. Despite appearances, however, semantic notions are rarely doing significant work in neuroscientific explanations; rather, computational descriptions are motivated primarily by other considerations. By contrast, representation often plays an ineliminable role in talk of computation in psychology. Consequently, talk of computation in neuroscience has little bearing on substantive psychological or philosophical theses about the role of computation in cognition. (Cao 2018: 283)

Of course, Cao could be wrong. Even so, the question arises of how literally to take talk of computation in neuroscience — about the ontological implications of computational explanations, one might say.

Consider a second example. In recent work, Mazviita Chirimuuta (2014, 201) challenges the conclusion that the prevalence of computational explanations in neuroscience requires the identification of brains with computers. One reason for Chirimuuta's doubt is her recognition that computational explanations in neuroscience are a variety of idealized explanation, similar to minimal causal models but also distinctive (Chirimuuta 2014; see also Weisberg 2007). Minimal causal models characterize their targets in terms of abstract and idealized causal relationships, where it may be false (i.e., an idealization) that the target literally has the features of the model, e.g., that a slope is in fact frictionless or a population infinite in size. Chirimuuta suggests that computational models in neuroscience can be understood as explaining the operation of the brain by characterizing it as implementing efficient coding procedures. She calls these, “i-minimal” models:

> Models which ignore biophysical specifics in order to describe the information processing capacity of a neuron or neuronal population. They figure in computational or information-theoretic explanations of why the neurons should behave in ways described by the model. (2014: 143)

Importantly, Chirimuuta denies that i-minimal models should be understood as describing the physical or neural mechanisms or causal processes that produce the phenomenon in question.

> We can now come to appreciate the distinctive explanatory value of I-minimal models. They figure prominently in explanations of *why* a particular neural system exhibits a particular empirically observed behaviour, by referring to its computational function.… Note that the appeal to coding principles like redundancy reduction does not involve decomposition of any mechanism thought to underlie the behaviour in question. Rather, it takes an observed behaviour and formulates an explanatory hypothesis about its functional utility. (2014: 144)

Chirimuuta is suggesting that computational explanations in neuroscience are not “abstract” explanations of the physical or neural causal processes in brain — they do not answer questions about *how* the phenomenon is produced. Instead, they give us a “computational” explanation in the sense advocated by David Marr in his famous three levels of explanation (1982). The computational explanation, in this sense, tells us what the system does and why. But it does not tell us how the phenomenon is produced – it does not provide details of the mechanism that "fills" the black box. If Chirimuuta is right, a computational explanation in neuroscience is not a literal description of the brain, because it is not a description of the brain at all. Instead, it is a redescription of the behavior of the system — the brain, in this case — in terms of the goals that it achieves. As the name connotes, an "i-minimal model" provides an interpretative framework rather than a literal description or characterization. In particular, it does not imply that the brain should be treated as literally or essentially a computer, or that brain processes are literally or essentially computational processes.

Elsewhere, Chirimuuta offers a slightly different but compatible proposal regarding the brain-as-computer hypothesis. She argues that this hypothesis should be understood as merely another analogy in a long history of such analogies, and no more ontologically committed than its predecessors (2021). She traces the idea that inference is a process of calculating as far back as Gottfried Leibniz, and cites him as an inspiration for more recent developments of the idea that brains implement reasoning as a computational process. But Chirimuuta argues that while the analogy between brains and computers may be useful, it ought not to be taken literally. And she is explicit that the brain-computer analogy should not be understood as itself implying that brains implement computational processes:

> My analogical interpretation will be presented as an alternative to the literal interpretations of neural-computational models which presume that the running of the model is a more or less accurate reproduction of a computation first instantiated in biological tissue. In order to pre-empt the worry that there is no substantial difference between the literal and analogical interpretations, I specify at the outset

> that I am not defining analogies as homomorphisms that obtain between the brain and its model. For on that definition the analogical relationship would amount to the instantiation of the same structure (i.e. function computed) in the neural system and the model. It would follow, on the assumption of a "mapping" account of computational implementation, that there would be no daylight between the literal and analogical interpretations of neurocomputational models, because the literal interpretation just is the claim that neural system and its model compute (approximately) the same function. (2021: 237)

What is Chirimuuta's reason for denying that the analogical view implies the kind of similarity that would count as computational implementation or realization? For her, an important reason is that analogical reasoning and understanding may essentially rely on the interests and capacities of those doing the explaining, whereas the literal interpretation of brains as computers, she assumes, must be free of such relativism or anthropocentrism: "The crucial point… is that the structure in the brain found to be relevantly similar to the model is not assumed to be an inherent, human-independent fact about the brain" (2021: 237).

In fact, among the advantages that Chirimuuta highlights for *not* taking the step from analogy to theory — for not taking computational explanations in neuroscience literally — is that the analogical view avoids vexed questions about multiple realization (Chirimuuta: 2021: 247-249).

We don't suppose that this brief summary of Chirimuuta's conclusions will convince many readers to give up the view that computational descriptions and explanations of cognition or brains should be taken literally. But it does, we think, suffice to illustrate that the literal interpretation is not the only option, even within a broadly realist explanatory context.

3. Computation and Abstractness

Suppose that there is some sense in which brain processes "really are" computational processes. Is it thereby obvious that brain processes or mental processes are multiply realizable?

The argument for a computational view of brains has these days shifted away from the contention that cognition should be explained in terms of computation and therefore that brains are realizing the cognitive "program" in us, i.e., away from the Computational Theory of Mind. More common now is the idea that brains themselves require explanation in terms of computation. That is, explanation in neuroscience, or some parts of neuroscience, is itself computational.

But here the argument from multiple realization runs into obstacles. For even if brains are essentially computing machines and therefore multiply realizable, it does not follow that *cognition* is essentially computational and therefore multiply realizable, as Putnam claimed.

Now, this might be thought of us a nitpicking point. If we think that the mind-brain identity theory says that mental states should be identified with brain states *and not with computational states*, then that would be falsified by the conclusion that brain states are, themselves, computational states. But that is just not the claim of the identity theory. The identity theory, even in its basic form, merely says, as J. J. C. Smart put it, "Sensations are brain processes" (1959). There's simply no further claim about the nature or function of these processes.

Of course this is not by any means a decisive argument against multiple realization. But at least we're reminding ourselves of what would have to be true for multiple realization to be possible. It would have to be true that cognition is literally a computing process — not merely analogous, or similar, or usefully thought of in that way — and that being computational in that sense is incompatible with the spirit if not the letter of the identity theory. Even so, would the fact that cognition is computational imply that cognition is multiply realizable? This question has special significance for those who see the autonomy of psychology as depending on the multiple realizability of cognition, which in turn they trace to the presumed computational nature of cognition.

Among philosophers, there is no received theory of the nature of computation that would enable us to univocally answer the question of whether computation is essentially multiply realizable. There are many theories of the nature of computation, with differing implications for the multiple realizability of computational processes.[6] This, by itself, ought to be enough to undermine the quick inference from being computational to being multiply realizable. But it doesn't. Why not?

One problem is a long-recognized but nevertheless typically ignored ambiguity in how philosophers and other specialists talk about the "abstractness" of computation (Lycan 1974; Van Gulick 1988; Putnam 1988; Cummins 1989; Chalmers 1996; Scheutz 1998, 1999; see also Maley forthcoming). The ambiguity surfaces when seeking to explain the sense in which computational states or processes are distinct from of their "hardware" implementations or realizers.

When Putnam introduces his hypothesis that pain states are functional states of the whole organism, explicated in terms of the idea of a probabilistic automaton, he uses scare quotes to describe the generalized automaton idea. For example, he writes:

> I shall assume the notion of a Probabilistic Automaton has been generalized to allow for 'sensory inputs' and 'motor outputs' — that is, the Machine Table

---

[6] For surveys, see Aizawa (2010), Miłkowski (2013a), and Piccinini (2015).

> specifies, for every possible combination of a ‘state’ and a complete set of ‘sensory inputs’, and ‘instruction’ which determines the probability of the next ‘state’, and also the probabilities of the ‘motor output’. (This replaces the idea of the Machine as printing on a tape.) I shall also assume that the physical realization of the sense organs responsible for the various inputs, and of the motor organs, is specified, but that the ‘states’ and the ‘inputs’ themselves are, as usual, specified only ‘implicitly’ — i.e., by the set of transition probabilities given by the Machine Table. (1967 in 1975: 433-434)

All the quote marks in the above passage are original to Putnam. They help him straddle the distinction between computing machines construed as abstract objects versus as physical systems. The resulting picture of functionalism rests on an ambiguity in the distinction between the “abstract” functional software and the physical hardware that implements the software. What is the “state” of the machine to which Putnam refers? Is it an essentially logical state, an abstractum? Or is it a state of a concrete entity, such as a physical machine, that is abstractly described, i.e., described with a subtraction of detail?

Computers as abstracta would have purely mathematical or logical properties and can have properties that no physical system may have. Famously, a so-called Turing Machine is stipulated to have infinite memory (Turing 1950). But there are no infinitely long recording tapes in the concrete world. Arguably, computational kinds considered as abstracta cannot be identified with physical kinds or any concrete kinds because they apparently have identity conditions that include possessing properties that no finite or concrete thing can literally instantiate. If so, no computational kind qua abstractum can be identified with any physical kind at all — just as no abstract mathematical or geometrical objects (if there are such things) can be identified with concrete physical objects. If this line of reasoning is cogent, then it is correct that we cannot identify computational kinds with physical kinds, and it could be true that computational kinds are inherently multiply realizable. But, in this story, the reason computational kinds cannot be identified with physical kinds does not itself depend on the multiple realizability of computation — it comes directly from computations being abstracta. This is why it is important not to equivocate when describing computational processes as abstract: they may be abstract in the manner of abstracta or in the more usual manner involving an "abstract" description that omits some physical details.

Suppose that computers as physical systems do not literally have mathematical, logical, or functional properties. That is, they do not *instantiate* properties of abstracta. Instead, the idea would be that rather than instantiating those sorts of properties, they realize or implement those properties by having some physical properties that stand in appropriate relations to one another *and to the abstract properties* (Cummins 1989; Polger 2004b, 2007). Typically the physical realization of abstract properties will be construed in terms of some sort of mapping or similarity relation, probably combined with additional constraints. For example, Putnam, noting that "an empirically given system can simultaneously be a 'physical realization' of many different Probabilistic Automata," defines the total functional state of a system relative to a description, S, of that system (1967 in 1975: 434). This should send up red flags, because whether some phenomenon is multiply realizable is supposed to be a fact about the world and not merely a fact about how the world is described.

The importance of the distinction between different notions of "abstractness" cannot be underestimated. If all there is to being a Probabilistic Automaton, or being a functional system more broadly, is to have "at least one Description of a certain kind" that maps the *physical* states of the system onto *abstract* machine states, then it does indeed seem obvious that infinitely or indefinitely many things could have such a description. This is the implication that Putnam himself celebrates, saying, "However, functional states of whole systems are something quite different. In particular, the functional-state hypothesis is *not* incompatible with dualism!" (1967 in 1975: 436, italics original).[7] That kind of computational function can be thought of as "abstract" in the sense that the physical system "realizes" or implements the logical, mathematical, and otherwise non-physical properties by standing in some external relation to those non-physical structures, e.g., a representational, mapping, or homomorphism relation (cf. Cummins 1989). Yet, on this view, no physical system literally has the abstract computational properties that it implements, for the abstract-as-abstracta computational properties are not themselves physical properties nor specified in physical terms.[8]

But there is an alternative view of computation that is more often applied to cognition and brains. Physical computing systems are not computers-as-abstracta. Physical computing states, transitions, and inputs or outputs are not abstract-qua-mathematical or logical but instead are concrete physical events and changes. Having certain concrete properties is not a matter of having "at least one Description of a certain kind" unless we highly restrict those descriptions to special classes of causal descriptions. Consequently, some physical computing systems might be highly constrained with respect to potential realizers, and others less so.[9] So even if it is correct that computers considered as abstracta could be realized by indefinitely or infinitely many different systems, that implication does not hold for physical computing. It's not at all clear that

---

[7] In contrast, Lewis (1972) coins the term "multiple realization" by way of expressing the concern that if the descriptions do not have unique realizers then they are "improper" descriptions that "do not (in any normal way) name anything here in our actual world" (1972: 252).

[8] If, as is common in recent literature, the realization relation is itself understood in terms of causal inheritance, as in the subset view (Wilson 1999, 2011) or dimensioned view (Gillett 2002), then abstract computational states are not even candidates for realization. Whether this is a defect in those accounts is a matter of dispute.

[9] This is closely related to what Lawrence Shapiro (2004) calls the Mental Constraint Thesis.

indefinitely or infinitely many things can realize every *physical* computing system (cf., Cao 2022, Maley forthcoming). The details would matter. And they would matter in a way that ought to undermine the intuition that physical computing machines are always multiply realizable. Physical computing systems are not always multiply realizable.

Then question, then, is: Should cognition be understood in terms of potentially constrained physical computing machines rather than obviously multiply realizable abstract computing machines? If the question is live — and it is — then that ought to be enough to temper the ready acceptance of the multiple realizability of cognition.

Putnam, himself, famously abandoned the abstract automaton theory, in part because he concluded that any physical system can realize any finite state automaton, and that the theory therefore fails to be a good theory of cognition (1988). Among current theorists of computation, there is disagreement between those who argue that the mapping notions of computation are not serious theories of computation at all (e.g., Piccinini 2015), and those who allow that they are theories of computation but argue that they are not the right notions of computation for explaining cognition (e.g., Milkowski 2013a, 2013b).

Even before Putnam rejected the theory, functionalists typically interpreted the theory in a way that replaces the idea of mental states as implicitly defined by a machine program with the idea that mental states are implicitly defined by an empirical or analytic theory, namely, a psychological theory (cf. Polger 2004). Significantly, these theories usually (though not always) define mental states in terms of causal roles, or causal functions. That is, they too leave behind the idea that the definitive features of cognition are purely abstract in the way that would make it obvious that they are multiply realizable qua abstracta. Causal relations, after all, are hostage to the properties of the physical materials that enter into these relations in the first place.

Moreover, current defenses of multiple realization in the brain tend to focus on neuroscience itself as confirming that brains are computational devices, and that those computational states are multiply realizable. Now, if you think that computational neuroscience shouldn't be taken too seriously (Bickle 2003) or shouldn't be taken too literally as making computational claims as suggested above (cf., Cao 2022, Chirimuuta 2021), then considerations about computational neuroscience will not carry weight for you. But we're trying to understand why many people think that multiple realization is obvious, so let's leave these reservations aside.

One of the clearest statements of the idea that brains are realizing physical computations and that those computations are multiply realizable comes from Gualtiero Piccinini, working with various collaborators. We will focus on his work as illustrative of the idea, while acknowledging the existence of alternatives.

Piccinini and Boone write, "Computational explanation is paradigmatically abstract" (2016: 1). Further, they say, "Some constitutive explanations are more *abstract* than others. That is, some constitutive explanations include fewer (causally relevant) details than others about their explanandum, their explanantia, or both" (2016: 3). This shows that the kind of abstraction that Piccinini and Boone take to be relevant to computational explanation in neuroscience is not of the abstract objects sort — not abstracta as contrasted with concreta. But it also relies on a notion of abstraction-as-subtraction that applies primarily to descriptions or representations and is not a feature of the things represented. The subtraction of detail does not occur at the object, but at the explanation. Things that are described abstractly, with detail left out, are not *ipso facto* multiply realizable. This is perhaps why this use of abstraction in explanations of other phenomena, such as the acceleration of balls rolling down inclined planes, never induces thoughts of multiple realization.

As noted above, Piccinini and Scarantino acknowledge that, “In many quarters, especially neuroscientific ones, the term ‘computation’ is used, more or less, for whatever internal processes explain cognition” (Piccinini and Scarantino 2010: 244). But they also report a specific use of the notion: “In recent decades, many neuroscientists have started using the term ‘computation’ for the processing of neuronal spike trains (i.e. sequences of spikes produced by neurons in real time). The processing of neuronal spike trains by neural systems is often called ‘neural computation’” (Piccinini and Scarantino 2010: 239).

And Piccinini formulates a theory of physical computation that is designed to include neural computation as a legitimate variety, even if it is not assimilated to digital or analog computation (2015). Appealing to this prior work, Piccinini and Scarantino write:

> We use ‘generic computation’ to designate any process whose function is to manipulate medium-independent vehicles according to a rule defined over the vehicles, where a medium-independent vehicle is such that all that matters for its processing are the differences between the values of different portions of the vehicle along a relevant dimension (as opposed to more specific physical properties, such as the vehicle’s material composition). Since a generic computation is medium-independent, it can be instantiated by many different physical mechanisms, so long as they possess a medium with an appropriate dimension of variation. (Piccinini and Scarantino 2010: 239)

The feature of Piccinini’s “generic computation” meant to guarantee multiple realization is the “medium-independence” of the computational vehicles.

In his most recent work, Piccinini defines medium independence in terms of multiple realizability. He says, “medium independence occurs when even the inputs and outputs of a higher-level property are multiply realizable” (2020: 66). Elsewhere, he explains,

> This feature of computers and their computing components — medium independence — is a stronger condition than MR [multiple realization]. Medium independence entails MR, because any medium independent property—such as performing a computation—can be realized by different kinds of medium. But MR does not entail medium independence, because typical multiply realizable properties—such as catching mice and lifting corks out of bottlenecks—can only be realized by mechanisms that manipulate the right kind of media—i.e., mice, corks, and bottles (Piccinini 2020: 41).

This way of understanding medium independence makes it a special case of multiple realization: medium independence involves multiply realizable properties that also have multiply realizable inputs and outputs. Piccinini is explicit on this point (2020: 62-66, esp. Fig. 2.4). But if we understand computation in terms of medium independence and we understand medium independence in terms of multiple realizability, then the line of reasoning from the mind being computational to it being multiple realizable will be question-begging. That is not what we want.

We need an independent account of the property of medium independence, and in fact Piccinini has an alternative explanation that does not presuppose multiple realizability. Piccinini and Scarantino say, “the functionally relevant aspects of neural processes are medium-independent aspects of the spikes—primarily, spike rates (as opposed to any more concrete properties of the spikes). Thus, spike trains appear to be another case of medium-independent vehicle, in which case they qualify as proper vehicles for generic computations” (Piccinini and

Scarantino 2010: 239). As they describe spike rates, they distinguish them from "concrete" properties of neural spikes, suggesting, by contrast, that they are abstract properties —properties of abstracta. But that is not what they intend. Remember, the relevant model of abstraction here is subtraction. Piccinini says:

> When we define concrete computations and the vehicles that they manipulate, we need not consider all of their specific physical properties. We may consider only the properties that are relevant to the computation, according to the rules that define the computation.… computational descriptions of concrete physical systems are sufficiently abstract as to be medium-independent. In other words, a vehicle is medium-independent just in case the rule (i.e., the input-output map) that defines a computation is sensitive only to differences between portions (i.e., spatiotemporal parts) of the vehicles along specific dimensions of variation—it is insensitive to any other physical properties of the vehicles. (Piccinini 2015: 122)

Similarly, Piccinini and Bahar write, "Computation in the generic sense is the processing of vehicles (defined as entities or variables that can change state) in accordance with rules that are sensitive to certain vehicle properties and, specifically, to differences between different portions (i.e., spatiotemporal parts) of the vehicles" (Piccinini and Bahar 2013: 458).

This avoids presupposing multiple realizability. Yet if the notion of "medium-independence" here is just that the processing of computational vehicles is sensitive only to *some but not all of their physical properties*, then this fails to distinguish computational explanation from other kinds of causal explanation.[10] And if computation is "medium-independent" only in that it depends on "differences between different portions (i.e., spatiotemporal parts) of the

---

[10] And that should be no surprise because Piccinini explicitly aims to assimilate computational explanation to mechanistic explanation (2015, 2020; see also Piccinini and Craver 2011).

vehicles" (Piccinini and Bahar 2013: 458), then computation is sensitive to the physical properties of the realizers, namely, to their spatiotemporal parts. If so, similar vehicles will have to be physically similar with respect to the relevant spatiotemporal parts. They might not be multiply realizable, after all. Medium independence in this manner, *pace* Piccinini (2020), is not a special case of multiple realization; it's just a different phenomenon altogether.

There is a connection between this critique of Piccinini's account of medium independence and our earlier discussion of the ambiguity between thinking of computing processes as abstracta versus their being describable abstractly, i.e., without full detail. As we noted, Piccinini and collaborators are clear that they intend abstraction in the sense of "subtraction of detail," but that comes into tension with their claim that computational properties (as opposed to descriptions) are abstract in the sense of medium independent. Piccinini writes, "If a description is sufficiently general, the property picked out by that description might have many lower-level realizers, but perhaps only trivially so; its multiple realizability might be an artifact of such a broad higher-level description that the property plays no useful role in scientific taxonomy or explanation" (2020: 42). But the multiple realizability of a property is not dependent on the completeness of its description.[11] If water is identical to H_2O, it does not become multiply realizable merely by being described as "the stuff that fills the oceans." In particular, subtraction of detail in a description does not make a property or state into an abstractum, so does not convert it into something for which there are few or no constraints on its realization.

In this section, we argued that even if computational explanations and models in the cognitive and brain sciences are to be taken as literally postulating that cognitive processes are

---

[11] Piccinini is aware of this issue (Piccinini and Boone 2016).

computational processes that are realized by brain processes, that does not imply that cognitive or brain processes are multiply realizable. In particular, once we look carefully at the phenomenon of "medium independence" that Piccinini argues is essential to computing, we see that it does not imply that all computational processes are multiply realizable. The reason is that the medium independence of computational processes is not a matter of their being a special kind of object, *abstracta* or *res calcula*, but just a way of thinking about or describing "abstractly" (i.e., without all detail) the physical systems that compute. Those may include brains and non-brains. But that does not imply that there is one thing, property, process, or kind that is literally present in each example and also differently realized in many examples. Hence, even if Piccinini and collaborators are right about the sense of computation that computational neuroscience adopts, that kind of computation does not imply multiple realizability.

Of course, on conceptions of computation other than Piccinini's, brains might be computing in a way that implies the ubiquitous multiple realizability of cognitive and brain processes. If so — and if there were good independent reasons for accepting this account of computation — we would be wrong about the prevalence of multiple realization in the cognitive and brain sciences. Despite the revision in our own view that this would require, it would mark an important insight, advancing our philosophical understanding of the cognitive and brain sciences. Yet even that "win" for multiple realization would be modest, because it would still not imply widespread multiple realization outside the brain sciences. Short of additional argument, it would not provide the basis for a general defense of the special sciences, as Putnam and Fodor argued it would.

4. The Argument Does Not Generalize

We've been considering why so many philosophers continue to believe that mental states are multiply realizable and therefore cannot be identified with brain states. The conjecture we're exploring is that the apparent obviousness of multiple realization is intertwined with the apparent transparency of the analogy between cognition and computing.

First, we suggested that insofar as cognition is analogous to computation, that alone might not ensure that brains are literally realizing computational processes. But suppose they are. In that case, we need to investigate whether the computational nature of cognition or brains is of a sort that promotes multiple realizability.

Second, we argued that the computational nature of cognition or brains need not entail that psychological kinds are multiply realizable. The reason is that the relevant sense in which cognition appears to be medium independent rests on a conception of abstraction-as-subtraction rather than abstraction-as-abstracta. But subtracting details from a description of cognitive processes no more makes them multiply realizable than does subtracting details from a description of water make *it* multiply realizable.

Now, suppose that's wrong and there is a multiple realization argument against the mind-brain identity theory. Another problem remains. Even if processing of medium-independent vehicles is the essential feature of computation and this implies multiple realization, thus falsifying the mind-brain identity theory, the multiple realization argument would not generalize in a way that Putnam and Fodor expected. It would not yield an account of why we have a plurality of sciences.

Why does the "multiple realization because of computation" argument not generalize? The reason is simple: not all special sciences explain by appeal to computation in the sense of processing of medium-independent vehicles. So, even if the reification of systems that manipulate medium-independent vehicles is crucial for the success of psychology or neuroscience, we have no reason at all to expect the same for anthropology, history, chemistry, meteorology, geology, and so on. At the very least, establishing that conclusion is well beyond what advocates of multiple realization have either argued or explicitly assumed.

5. The Plurality of Sciences

Putnam and Fodor argued that we have special sciences because of multiple realization. We have special sciences — we have a plurality of sciences rather than just one — because the natural kinds of the special sciences are not identifiable with the natural kinds of physical science. As Fodor, in his classic paper, "The Special Sciences, of the Disunity of Science as a Working Hypothesis," wrote:

> The reason it is unlikely that every natural kind corresponds to a physical natural kind is just that (a) interesting generalizations (e.g., counter-factual supporting generalizations) can often be made about events whose physical descriptions have nothing in common, (b) it is often the case that whether the physical descriptions of the events subsumed by these generalizations have anything in common is, in an obvious sense, entirely irrelevant to the truth of the generalizations, or to their interestingness, or to their degree of confirmation or, indeed, to any of their epistemologically important properties, and (c) the special sciences are very much in the business of making generalizations of this kind. (1974: 103)

As many have noted, Fodor does not offer a detailed argument for any of these claims (e.g., Jones 2004). Supporting his confidence is simply more confidence: “I take it that these remarks are obvious to the point of self-certification; they leap to the eye as soon as one makes the (apparently radical) move of taking the special sciences at all seriously” (1974: 103).

What argument Fodor provides is by way of an example:

> Suppose, for example, that Gresham's ‘law’ really is true. (If one doesn't like Gresham’s law, then any true generalization of any conceivable future economics will probably do as well.) Gresham's law says something about what will happen in monetary exchanges under certain conditions. I am willing to believe that physics is general in the sense that it implies that any event which consists of a monetary exchange (hence any event which falls under Gresham’s law) has a true description in the vocabulary of physics and in virtue of which it falls under the laws of physics. But banal considerations suggest that a description which covers all such events must be wildly disjunctive. Some monetary exchanges involve strings of wampum. Some involve dollar bills. And some involve signing one's name to a check. What are the chances that a disjunction of physical predicates which covers all these events (i.e., a disjunctive predicate which can form the fight hand side of a bridge law of the form ‘x is a monetary exchange [if and only if]…’) expresses a physical natural kind? In particular, what are the chances that such a predicate forms the antecedent or consequent of some proper law of physics? The point is that monetary exchanges have interesting things in common; Gresham’s law, if true, says what one of these interesting things is.” (Fodor 1974: 103)

Fodor goes on to conclude, “A natural kind like a monetary exchange could turn out to be co-extensive with a physical natural kind; but if it did, that would be an accident on a cosmic scale” (Fodor 1974: 104).

Of course, it’s a familiar fact that different tokens can be used for economic transactions. The question is whether that fact implies that there is a single natural kind of process, viz., monetary exchanges, that is multiply realized. And, if so, is it plausible that minds and hearts are multiply realized in the same way, so that the example lends general support to the legitimacy and irreducibility of special sciences across the board, as Fodor claims?

We think that Fodor’s example depends on an ambiguity about “monetary exchange” akin to the ambiguity about computation we discussed above. Fodor seems to have in mind financial transactions as exchanges of abstract quantities, like numbers. Anything at all could, in principle, be made to stand for those quantities or values, so that exchanges of those tokens, whatever they may be, would count as financial transaction. But his example does not support that way of thinking.

In monetary exchanges, it’s not true that all of the choices of tokens are equally good — as Fodor’s own example of Gresham’s Law illustrates. Gresham’s Law says that “bad money drives out good.” What does that mean? Well, it means that some currencies are favored by investors because they are more likely to retain or grow their value compared to others. Many countries and individuals invest in US dollars because they are often considered more reliable in the long term than other currencies because they are guaranteed by the US government. But the origins of Gresham’s Law are even simpler and more concrete. If coins, such as gold and silver coins, are supposed to have the same commodity value as their face value, then they need to have certain stable physical features. In particular, their weight must remain unchanged, as it is the basis for the commodity value. But as coins get older, they wear down. Some of their commodity value is lost, and they are no longer worth their face value. Or some unscrupulous person might shave a tiny amount of metal off each coin, undetectable to the naked eye, but adding up over many coins to a substantial deficit. This is “bad” money, money that has a commodity value that is lower than its face value. Given the choice, people will tend to spend bad money and save “good” money, that is, save the money that has the correct commodity value. In a marketplace where bad money is present, people will tend to hold on to their good money and spend their bad money, so the money in exchange will be mainly bad money — the good money will be driven out of the marketplace and into savings. Bad money drives out good.

Notice that, far from being an abstract (qua abstracta) principle about idealized financial exchanges, Gresham’s Law is exactly about why the vehicles of financial transaction sometimes matter. Fodor’s mistake was to think that just because there is a very general principle, one that applies to various different kinds of currencies, that it was therefore a principle or law about a special kind of event — one that could abstract away from the physical details of monetary exchanges. Yet it is not. It is an entirely realization-dependent principle, albeit one that can occur in a variety of ways. But the lack of detail in the principle does not imply that the objects it applies to are of a special sort, computational or abstract.[12] Put in the terms discussed above, the apparent medium independence of financial transactions does not imply that they are exchanges of abstract objects or properties, nor that there are few or no constraints on their vehicles.

You might think that the other classic example of multiple realization does not share this defect. Hilary Putnam (1972) argued that the right explanation of why a rigid square peg does

---

[12] Notice that the economic example also has the feature emphasized by Chirimuuta (2021): what counts as a “monetary exchange” or “financial transaction” is dependent on human beings and our interests. The fact that anything could be used as a currency is just the fact that anything can be used as a conventional or arbitrary representation — that is, used by us to represent, given the right circumstances.

not pass through a round hole in a rigid board is in terms of the macroscopic shape of the peg and hole, regardless of what they are made of, and not in terms of the particular molecular structure of the objects and their trajectories over time as predicted by physical laws: “The peg passes through the hole that is large enough to take its cross-section, and does not pass through the hole that is too small to take its cross-section. That is the correct explanation whether the peg consists of molecules, or continuous rigid substance, or whatever” (Putnam 1975: 296).

Putnam leaves us with the impression that the shape of the entire board or peg is a property of the macroscopic thing and not the microscopic, and that is why the apparently macroscopic explanation that mentions pegs, holes, and their “structural features” is to be preferred over the microscopic explanation. He says, “The same explanation will go in any world (whatever the microstructure) in which those higher level structural features are present. In that sense this explanation is autonomous” (1975: 296). And, he continues, “People have argued that I am wrong to say that the microstructural deduction is not an explanation. I think that in terms of the *purposes for which we use the notion of explanation*, it is not an explanation. If you want to, let us say that the deduction *is* an explanation, it is just a terrible explanation, and why look for terrible explanations when good ones are available?” (1975: 296)

But are the properties that Putnam’s preferred “macro” explanation invokes really physical properties of the square peg and round hole? On reflection, it’s not so clear. Putnam says, “The explanation is, the board is rigid, the peg is rigid, and as a matter of geometrical fact, the round hole is smaller than the peg, the square hole is bigger than the cross-section of the peg” (1975: 296). And, after asserting that the board and peg could be made of anything at all, he writes, “If one wanted to amplify the explanation, one might point out the geometrical fact that a square one inch high is bigger than a circle one inch across” (1975: 296).

When we pay attention to the fact that these “higher level” properties are geometrical properties, it seems that what gives the explanation its distinctive autonomy is not that pegs and boards are multiply realizable, or could be made of anything at all, but rather that the explanation is fundamentally not a physical explanation but a geometrical one. Why does the square peg not fit through a round hole that has the same diameter? Well, because it’s impossible to inscribe a square into a circle whose diameter is the same as the length of the square’s sides; and this situation exhibits those geometrical features.

Of course there is a physical aspect to the explanation: the objects are assumed to be rigid solids. The peg can very well pass through the hole if it can change shape, if the “hole” is made of air, and so on. So it’s an interesting contingent fact that these items approximate the geometrical conditions as they do. But this defense merely creates a dilemma for Putnam. The explanation for the peg’s failure to fit through the round hole must either refer to the physical properties of the peg and the board in which the hole is carved, or else refer to abstract geometric features of squares and circles. In the first case, physical details do matter: a peg constructed from malleable material will fit through the hole. The explanation, in other words, is medium dependent. In the second case, which Putnam seems to favor, the explanation seems to rely mainly on the geometrical features, and therefore borrows the “abstractness” of geometry as concerning abstract objects, ignoring that the assumed physical background conditions — rigidity, solidity, and so on — may indeed provide the basis for a physical commonality of the physical conditions under which the geometrical generalization may be usefully applied. The reason that “*the ultimate constituents don’t matter*, that *only the higher level structure matters*” (Putnam 1975: 296) is simply that the relevant generalization is not a physical law but a geometrical theorem. That is, it's not a special sciences law that covers multiply realized kinds. It's a mathematical explanation.[13]

Seen this way, Putnam's example of the square peg and the round hole looks very much like Marc Lange's example of dividing strawberries. Why can a mother with 3 children not divide 23 strawberries equally among her children without cutting them? Lange states the obvious:

> The fact that 23 cannot be divided evenly by 3 explains why Mother fails every time she tries to distribute exactly 23 strawberries evenly among her 3 children without cutting any (strawberries— or children!). (2016: 6)

Notice that Lange's example has the distinctive feature cited by Putnam, Fodor, and other advocates of multiple realization: It doesn't matter what the things being divided are or what they made of. They could be grapes, melons, popsicles, spiders, cars, houses, or planets. The explanation is the same.

But that does not imply that there is a single kind, such that it cannot be evenly divided into 3 without cutting, that is multiply realizable by grapes, melons, popsicles, spiders, cars, houses, planets, and indefinitely other kinds of things. This situation does not call for a physical explanation that covers multiply realized kinds. It calls for a different kind of explanation. Lange argues that it calls for a mathematical explanation:

> the explanation of Mother's failure to distribute her 23 strawberries evenly (without cutting any) among her 3 children reveals that her failure is insensitive to her precise technique for distributing strawberries. However, this explanation does not work by describing the various causal relations that would have obtained

---

[13] Compare the debate over whether the Principle of Natural Selection is an empirical law or a trivial theorem of probability theory (e.g., Brandon 1990).

> if she had tried other ways of distributing the strawberries, and then showing that each of these causal processes would have failed to distribute the strawberries evenly. This explanation does not describe any technique of strawberry distribution. It derives its explanatory power neither by virtue of describing the actual causes nor by telling us about what the network of causal relations would have been like under other initial conditions. Rather (I will argue), this explanation shows the outcome to be inevitable to a stronger extent than facts merely about causal relations (actual and counterfactual) could make it. This explanation is distinctively mathematical. (2016: 14-15)

The present point is not to secure Lange's conclusion, or to assume its correctness. At this point, our objective is merely to point out the similarity between Lange's strawberries example, Putnam's square peg and round hole, and Fodor's monetary exchanges. It is arguable, in each case, that the actual or apparent explanatory generality at stake does not come from the existence of a special science law that covers kinds that are "abstract" and thereby multiply realizable, but instead from genuinely or apparent abstract mathematical explanations that apply to a wide variety of things whether or not they realize some common kind. Putnam seems to think of his square peg explanation as physical, but it appears to be more like a mathematical explanation. Fodor seems to think of economic explanations as more like a mathematical explanations in their abstractness, but Gresham's Law is fully medium dependent.

6. Conclusion

A common reason for thinking that multiple realization is ubiquitous or obvious derives from the view that cognition is computational, or that brains are computing machines. These views, in turn, receive support from the apparent prevalence of computational explanations in the cognitive and brain sciences. And it has seemed obvious to many that whatever is computational is multiply realizable.

We've argued that the mere presence of computational explanations in the cognitive and brain sciences does not justify the conclusion that multiple realization is ubiquitous or obvious. One reason is that computational explanations in the cognitive and brain sciences should not always be taken literally. The second reason is that even if we take computational explanations in the cognitive and brain sciences literally, we have to distinguish different notions of computation. If we think of computation as involving relations among abstracta, they might appear to be multiply realizable, but that is not how we should think about computing in the brain. If we think in terms of physical computation, we still lack grounds for multiple realizability. For example, we argued that Piccinini's view of neural computation as manipulating medium independent vehicles does not imply multiple realizability of all computing processes. Finally, we argued that even if it turns out that cognitive and brain processes are multiply realizable because they are computational in some suitable sense, the argument based on computation does not generalize to support the autonomy and plurality of sciences outside of the cognitive sciences, such as chemistry, anthropology, meteorology, economics, and so on.

At this point it's worth highlighting a paradoxical feature of the Putnam-Fodor argument for the autonomy of the special sciences. Putnam says, "the ultimate constituents don't matter" (1975: 296, italics removed). Fodor says, "whether the physical descriptions of the events subsumed by these generalizations have anything in common is, in an obvious sense, entirely irrelevant to the truth of the generalizations" (1974: 103). We think that's right. It doesn't matter if the kinds of special sciences have something physically in common. But to then argue that special science explanations get their autonomy because they cover kinds that are multiply realized is a confusion. If it doesn't matter whether the things covered by special science explanations have something physically in common, then it doesn't matter that they don't have anything in common, either.

This is a point raised by Polger and Shapiro, who argue that the "autonomy" or legitimacy of the special sciences does not depend on the truth of multiple realization, after all (2016). If the plurality of sciences does not depend on multiple realization, then what grounds their legitimacy and autonomy?

That is an important question, one version of the question, "Why is there anything but physics?" Polger and Shapiro's answer appeals to a rejection of explanatory exclusion principles that will strike many philosophers as requiring more justification than they offer. But at least it has the virtue of not resting the case for the special sciences on the disputed fact of multiple realization.

Conflict of Interest Statement

The authors have no relevant financial or non-financial interests to disclose.

References

Aizawa, K. 2010. Computation in cognitive science: it is not all about Turing-equivalent computation. *Studies in History and Philosophy of Science* 41: 227–236

Aizawa, K. and C. Gillett. 2009a. The (Multiple) Realization of Psychological and Other Properties in the Sciences. *Mind and Language*, 24: 181–208.

Aizawa, K. and C. Gillett. 2009b. Levels, Individual Variation, and Massive Multiple Realization in Neurobiology. In J. Bickle (Ed.), *Oxford Handbook of Philosophy and Neuroscience*. New York: Oxford University Press: 529–581.

Aizawa, K. and C. Gillett. 2011. The Autonomy of Psychology in the Age of Neuroscience. In Phyllis McKay Illari Federica Russo (ed.), *Causality in the Sciences*. Oxford University Press: 202-223.

Appiah, K. A. 2017. *As If*. Cambridge, MA: Harvard University Press.

Bechtel, W. and J. Mundale. 1999. Multiple Realizability Revisited: Linking Cognitive and Neural States. *Philosophy of Science*, 66: 175-207.

Bickle, J. 1998. *Psychoneural Reduction*. Cambridge, MA: The MIT Press.

Bickle, J. 2003. *Philosophy and Neuroscience: A Ruthlessly Reductive Account*. Kluwer Academic Publishers, Boston, MA.

Bickle, J. 2010. Has the Last Decade of Challenges to the Multiple Realization Argument Provided Aid and Comfort to Psychoneural Reductionists? *Synthese* 177: 247–260.

Bickle, J. 2020. Multiple Realizability. *The Stanford Encyclopedia of Philosophy* (Summer 2020 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/sum2020/entries/multiple-realizability/>.

Block, N. and J. Fodor. 1972. What Psychological States Are Not. *Philosophical Review* 8: 159-81.

Block, N. 1997. Antireductionism Slaps Back. Philosophical Perspectives, 11: 107-132.

Brandon, R. 1990. *Adaptation and Environment*. Princeton: Princeton University Press.

Cao, R. 2018. Computational Explanations and Neural Coding. In Sprevak and Colombo (Eds), *Routledge Handbook of the Computational Brain*: New York: Routledge, 2018: 283-296.

Cao, R. 2022. Multiple Realizability and the Spirit of Functionalism. *Synthese*, 200 (6):1-31..

Chalmers, D. 2022. *Philosophy of Mind: Classical and Contemporary Readings* (Second Edition). New York: Oxford University Press.

Chalmers, D. J. 1996. Does a Rock Implement Every Finite-State Automaton? *Synthese* 108: 310–333.

Chirimuuta, M. 2014. Minimal Models and Canonical Neural Computations: The Distinctness of Computational Explanation in Neuroscience. *Synthese*191 (2): 127-153.

Chirimuuta, M. 2021. Your Brain Is Like a Computer: Function, Analogy, Simplification. In F. Calzavarini, M. Viola (eds.), *Neural Mechanisms, Studies in Brain and Mind* 17: 235-260.

Craver, C. 2004. Dissociable Realization and Kind Splitting. *Philosophy of Science*, 71: 960-971.

Cummins, R. 1989. *Meaning and Mental Representation*. Cambridge MA: MIT Press.

Dennett, D. 1983. "Styles of Mental Representation," *Proceedings of the Aristotelian Society* 83: 213-226.

Enç, B. 1983. In Defense of the Identity Theory. *Journal of Philosophy* 80: 279-98.

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. In Tomberlin, J. (ed). *Philosophical Perspectives 11: Mind, Causation, and World*. Boston: Blackwell Publishers (1997).

Gillett, C. 2002. The Dimensions of Realization: A Critique of the Standard View. *Analysis* 62 (4): 316-323.

Hatfield, G. 1991. “Representation in Perception and Cognition: Connectionist Affordances,” in *Philosophy and Connectionist Theory,* eds. W. Ramsey, S. Stick and D. Rummelhart (Hillsdale, NJ: Lawrence Erlbaum Associates, 163-195.

Hemmo, M. and O. Shenker. 2015. The emergence of macroscopic regularity. Mind and Society 14 (2): 221-244.

Jones, T. 2004. Special Sciences: still a flawed argument after all these years. *Cognitive Science*, 28: 409–432.

Kim, J. 1972. Phenomenal Properties, Psychophysical Laws, and Identity Theory. *Monist* 56, 2: 177-192.

Kim, J. 1993. *Supervenience and Mind: Selected Philosophical Essays*. Cambridge, UK: Cambridge University Press.

Kim, J. 2011. *Philosophy of Mind* (Third Edition). Philadelphia, PA: Westview Press.

Kitcher, P. 1984. 1953 And All That: A Tale of Two Sciences. *Philosophical Review* 93 (3): 335-373.

Kosslyn, S. and Hatfield, G. 1984. “Representation without Symbol Systems,” *Social Research* 51: 1019-1045.

Lange, M. 2016. *Because Without Cause: Non-Causal Explanations in Science and Mathematics*. Oxford. UK: Oxford University Press.

Levin, J. 2008. Functionalism. *The Stanford Encyclopedia of Philosophy* (Winter 2021 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/win2021/entries/functionalism/>.

Lewis, D. 1969. Review of Art, Mind, and Religion. *Journal of Philosophy* 66: 23-35.

Lewis, D. 1972. Psychophysical and Theoretical Identifications. *The Australasian Journal of Philosophy* 50:249-258.

Lycan, W. 1974. Mental States and Putnam's Functionalist Hypothesis. *Australasian Journal of Philosophy* 52: 48-62.

Maimon, A. and M. Hemmo. 2022. Does Neuroplasticity Support the Hypothesis of Multiple Realizability? *Philosophy of Science*, 8: 107–127

Maley, C. forthcoming. Medium Independence and the Failure of the Mechanistic Account of Computation. *Ergo.*

Marr, D. 1982. *Vision*. Cambridge, MA: The MIT Press.

Miłkowski, M. 2013a. *Explaining the Computational Mind*. Cambridge, MA: The MIT Press.

Milkowski, M. 2013b. A Mechanistic Account of Computational Explanation in Cognitive Science. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Cooperative Minds: Social Interaction and Group Dynamics*. *Proceedings of the 35th Annual Meeting of the Cognitive Science Society* (Austin, TX: Cognitive Science Society): 3050–3055.

Piccinini, G. 2015. *Physical Computation*. Oxford: Oxford University Press.

Piccinini, G. 2020. *Neurocognitive Mechanisms: Explaining Biological Cognition*. New York: Oxford University Press.

Piccinini, G. and A. Scarantino. 2011. Information Processing, Computation, and Cognition. *Journal of Biological Physics* 37 (1):1-38.

Piccinini, G. and C. Craver. 2011. Integrating Psychology and Neuroscience: Functional Analyses as Mechanism Sketches. *Synthese* 183: 283–311.

Piccinini, G. and O. Shagrir. 2014. Foundations of Computational Neuroscience. *Current Opinion in Neurobiology* 25 (2014): 25-30.

Piccinini, G. and S. Bahar. 2013. Neural Computation and the Computational Theory of Cognition. *Cognitive Science* 37 (3): 453-488.

Piccinini, G. and T. Boone. 2016. Mechanistic Abstraction. *Philosophy of Science*, 83: 5: 686–697.

Polger, T. 2002. Putnam’s Intuition. *Philosophical Studies*, 109, 2: 143-170.

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

Polger, T. 2004b. Neural Machinery and Realization. *Philosophy of Science*. 71 (5): 997-1006.

Polger, T. 2007a. Realization and the Metaphysics of Mind. *Australasian Journal of Philosophy*, 85 (2): 233-259.

Polger, T. 2009a. Evaluating the Evidence for Multiple Realization. *Synthese* 167 (3): 457-472.

Polger, T. 2009b. Two Confusions Concerning Multiple Realizability. *Philosophy of Science* 75 (5): 537-547.

Polger, T. and L. Shapiro. 2016. *The Multiple Realization Book*. Oxford, UK: Oxford University Press.

Putnam, H. 1967. Psychological Predicates. Reprinted as “The Nature of Mental States” in H. Putnam (ed.) 1975.

Putnam, H. 1973. Philosophy and Our Mental Life. Reprinted in H. Putnam (ed.) 1975.

Putnam, H. 1975. *Mind, Language, and Reality: Philosophical Papers, Volume 2*. Cambridge: Cambridge University Press.

Putnam, H. 1988. *Representation and Reality.* Cambridge MA: MIT Press.

Richardson, 2009. Multiple Realization and Methodological Pluralism. *Synthese* 167 (3): 473 - 492.

Richardson, R. 1979. Functionalism and Reductionism. *Philosophy of Science*, 46: 533-558.

Scheutz, M. 1998. Implementation: Computationalism's Weak Spot. *Conceptus* 31: 229-39.

Scheutz, M. 1999. When Physical Systems Realize Functions. *Minds and Machines* 9: 161-96.

Shapiro, L. 2000. Multiple Realizations. *Journal of Philosophy* 97 (12):635-654.

Shapiro, L. 2004. *The Mind Incarnate*. Cambridge, MA: The MIT Press.

Shapiro, L. 2008. How to Test for Multiple Realization. *Philosophy of Science* 75 (5):514-525.

Shapiro, L. 2010. Lessons From Causal Exclusion. *Philosophy and Phenomenological Research*, 81: 594–604.

Smart J. J. C. 1959. Sensations and Brain Processes. *Philosophical Review*, LXVIII: 141-156.

Sober, E. 1999. The Multiple Realizability Argument Against Reductionism. *Philosophy of Science* 66: 542-564.

Thagard, P. 2022. Energy Requirements Undermine Substrate Independence and Mind-Body Functionalism. *Philosophy of Science*, 89: 70–88.

Turing, A. 1950. Computing Machinery and Intelligence. *Mind*, Volume LIX, Issue 236: 433–460.

Van Gulick, R. 1988. Consciousness, Intrinsic Intentionality, and Self-Understanding Machines, in *Consciousness in Contemporary Science*, ed. A. J. Marcel and E. Bisiach, (Oxford University Press: Oxford) 78-100.

Weisberg, M. 2007. Three Kinds of Idealization. *Journal of Philosophy* 104 (12):639-659.

Wilson, J. 1999. How Superduper Does a Physicalist Supervenience Need To Be? *The Philosophical Quarterly* 49: 33-52.

Wilson, J. 2011. Non-reductive Realization and the Powers-based Subset Strategy. *The Monist.* Vol. 94, No. 1, Powers (January 2011), pp. 121-154.