"Marr, Mayr, and MR: What Functionalism Should Now Be About"

M. Chirimuuta

History & Philosophy of Science, University of Pittsburgh

Final version: 2018 *Philosophical Psychology,* 31(3):403-418

Abstract

In this review essay on *The Multiple Realization Book* by Polger and Shapiro (Oxford University Press, 2016), I consider the prospects for a biologically grounded notion of multiple realization which has been given too little consideration in the philosophy of mind and cognitive science. Thinking about MR in the context of biological notions of function and robustness leads to a rethink of what would count as a viable functionalist theory of mind. I also discuss points of tension between Polger and Shapiro's definition of MR and current explanatory practice in neuroscience.

1. MR 1.0

The thesis that mental states are multiply realized by neural states, and the functionalist theory of the mind given credence by it, are two of the rarest of beasts in philosophy: ideas that a near majority of philosophers have subscribed to, and for more than one generation. The presumption that the existence of multiple realization is an empirically established fact is part of the explanation of the surprising longevity of a non-reductive physicalist consensus. In *The Multiple Realization Book,* Thomas Polger and Lawrence Shapiro challenge that presumption head on with an impressive deployment of arguments and extensive review of scientific facts. Without doubt, the MR debate will not be the same again.

Polger and Shapiro's wide-ranging critical project is intended to undermine the entrenched anti-reductive consensus in order to clear the ground for reductive theories to thrive (p.12). The purpose of my essay is to complicate their vision of the post-functionalist future. I, for one, have been convinced by Polger and Shapiro that the original conception of multiple realization ("MR 1.0") needs to go -- they have done a service to philosophy by showing the weaknesses of the received view, and written a sophisticated and rewarding book in the process. However, as I will argue below, we still need MR. This because of the centrality of functional thinking *within* biology and neuroscience – a topic that P&S do not delve into. It is time to imagine a new conception of ("MR 2.0"), which I will sketch out below. First we must discuss MR 1.0 and the reasons why its time is up.

i. Conditions for Multiple Realization

In this brief characterisation of the original MR thesis I draw on Putnam (1967; 1973) and Fodor (1974) rather than the many later publications on the subject. On this way of framing things, the question "is mind reducible to brain?" becomes "is psychology reducible to neuroscience?". Psychology deals with functions, neuroscience with their realizers. The Fodor-Putnam account comes with the problematic implication that neuroscience is not relevant to understanding the mind. A functionalist metaphysics of mind also leads Fodor and Putnam to bet on there being widespread and dramatic instances of multiple realization in nature – if a mind, fundamentally, is the a kind of thing that *could* have been made out of anything (Swiss cheese included), then we should expect millions of years of evolutionary experiment to have explored this possibility space and thrown up at least a few surprisingly different realizations.

Much of Polger and Shapiro's careful survey of the neuroscientific literature (Chapters 5 and 6) serves to make a compelling case that this prediction has not been borne out by the actual facts. They utilize their own characterisation of MR that is developed and defended in Chapters 3 and 4 (the "official recipe"). Here are the conditions that a comparison of kinds from two different sciences must satisfy in order to qualify as a case of multiple realization. This list of requirements is the beating heart (or pulsating brain) of the book:

i. As and Bs are of the same kind in model or taxonomic system S1.
ii. As and Bs are of different kinds in model or taxonomic system S2
iii. 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
iv. The relevant S2-variation between As and Bs must be distinct from the intra-kind variation between As and Bs. (p. 67)

Condition *iii* ensures that the candidates for different realizations differ in ways which are relevant to the function that they are both said to perform. For illustrative purposes P&S return to their old example of the waiter's and two handled corkscrews. The difference in basic design plan (e.g. one vs. two pivots) leads to the different S2 classification, while the fact that the properties referred to in S2 classification bring about the shared S1 functional classification of removing corks means that these are genuinely different realizations of the same function. One may worry about reliance on these simple artefacts as the template for MR in formidably complex and only partially understood biological systems. I will return to this concern below.

Variants of different material constitution do not automatically count as multiple realizations because of condition *iv*. Plastic and wooden handled corkscrews will differ in strength and weight just as octopus and human eyes differ in their focussing ability (p.42), and human retinas show variation in light capturing photo-pigments which they contain (p.104-111). None of these kinds of variations multiply realise the function of the eye because they all come within the scope of normal within-kind variation. And on the assertion that human and octopus eyes are mere variants of the same camera eye design, it turns out Putnam (1967) would agree.

Below in Section 3(iii) I will consider whether or not these conditions for MR are too demanding. Here I will raise a broader point about framing. Polger and Shapiro follow Fodor (1974) closely in taking the issue at stake to be about the cross-classification of the kinds of different branches of science, restricting their attention narrowly on the functionalist's conception of MR in philosophy of mind. This is all well and good if the only aim is to score the dialectical point against functionalism. However, there are some problematic features of MR 1.0 which themselves infect Polger and Shapiro's treatment, and limit the usefulness of their analysis to broader projects within the philosophy of the mind-brain (e.g. developing accounts of explanation in psychology and neuroscience). A major failing of the MR 1.0 framework is that it does not recognise the importance of functional thinking *within* biology (and hence neuroscience), and moreover it pits those who think biology and neuroscience are relevant to understanding the mind against those with a non-reductionist bent. As will become clear in the course of this essay, this is a troubling dichotomy.

2. Marr -- and Others -- on Functional Explanation in Neuro- and Cognitive Science

Discussions of MR and functionalism within cognitive science itself often bring us to David Marr's three levels. Although he refers them as levels of "description" or "understanding" (Marr 1982:24-25), they are often treated by philosophers as levels of *being*, perhaps under the influence of Fodor's (1974) discussion of 'kinds' and 'properties' of the physical and special sciences, and laws governing them. On this Fodorian picture, computational theory describes psychological kinds, the 'mental', and its functions; implementation theory describes neural kinds, the 'physical' – whatever realizes psychological functions. This picture is consistent with Polger and Shapiro's definition of MR as a relationship between kinds of different sciences.[1]

Marr is often erroneously bracketed with Fodor as an advocate of the full autonomy of psychology from neuroscience.[2] Thus advocates of computational theory and functional description are pitted against those who see the relevance of brain science to understanding the mind. However, this invidious choice between acknowledging the value either of computational approaches or neural ones is eliminated as long as we take Marr at his word as outlining different

[1] See Boone (forthcoming) for criticism of this approach, and an alternative causal framework.
[2] For further discussion see Kaplan (2011:342-343) and Chirimuuta (forthcoming).

types of explanation of one thing – the visual system. Computational explanation occurs within neuroscience, not just psychology;[3] neuroscience happens also to be in the business of discovering mechanisms which implement the proposed functions.

As the vision scientists Frisby & Stone (2012:1042) write:

> "Some take the view that finding a link between a given visual phenomenon and a neurophysiological process counts as a sufficient explanation of that phenomenon. We disagree, as did Marr ….. A theory of vision should do more than identify a mechanism that could implement a given computation: it should also provide a functional (e.g. computational) reason why the computation was desirable in the first place. To some extent, Marr's call for a computational account of vision has been taken up in the nascent field of computational neuroscience, where fine-grained analysis of physiological data is commonly interpreted in terms of its functional significance."

These authors point out the insufficiency of purely mechanistic explanation – ones that do no more than identify the neural processes underlying visual phenomena. Their appeal is just to what they find to be a satisfying explanation. However, there are in-principle reasons why explanatory pluralism is to be expected in neuroscience. As Marr (1982) noted, and philosophers of biology like Mitchell (2009) have also argued, highly complex biological systems require description and explanation with multiple kinds of models and theoretical approaches. Each of these perspectives guides researchers to certain phenomena, and suggests which details of the system might be safely ignored.

[3] This is an obvious point, and it has been true since the 1950's; but it is neglected by functionalists in the Fodor-Putnam tradition. Putnam (1973) himself recognises the importance of levels of explanation, giving the square peg/round hole example. Putnam says that the higher level explanation is better because it is "far more general" and at one point the autonomy of the mental amounts just to the credible claim that, "[w]hatever our mental functioning may be, there seems to be no serious reason to believe that it is explainable by our physics and chemistry". Yet this sensible discussion of explanation is part of the larger argument that brain material is irrelevant to explaining mental life.

Note that on this perspectival and pluralist account the computational description is no more and no less realistic and fundamental than the mechanistic/implementational one.[4] Yet, computational theory does have a preeminent role to play in cutting through the daunting complexity of neural structures and activity patterns. This was recognised early in the history of visual neurophysiology. The following passage comes from an article in which Horace Barlow presents the redundancy reduction explanation for the existence of lateral inhibition in the retina, deriving the result from information theory:[5]

> "A wing would be a most mystifying structure if one did not know that birds flew. . . . [W]ithout understanding something of the principles of flight, a more detailed examination of the wing itself would probably be unrewarding. I think that we may be at an analogous point in our understanding of the sensory side of the central nervous system. We have got our first batch of facts from the anatomical, neurophysiological, and psychophysical study of sensation and perception, and now we need ideas about what operations are performed by the various structures we have examined. . . .
>
> It seems to me vitally important to have in mind possible answers to this question when investigating these structures, for if one does not one will get lost in a mass of irrelevant detail and fail to make the crucial observations." (Barlow, 1961:217)

The idea here is that knowledge of the function of a structure of process leads to knowledge of the relevant theoretical principles that apply to it -- aerodynamics for flight, information theory for the nervous system. With the appropriate theoretical framework in place, the scientist can better see what the relevant details are for explaining the behaviour of the system, and this is crucial for scientific progress given the over-production of data. The point is not about the

[4] The way to argue otherwise would be via a blanket argument for reductive physicalism which prioritises the 'lower level' description. But then why stop at neural description, rather than chemical, then physical?

[5] See Chirimuuta (2017a) for an extended discussion.

intuitive insufficiency of purely mechanistic explanation (as with Frisby and Stone, quoted above); rather, it is about how to make scientific progress given the complexity of the brain.[6]

3. Mayr -- and Other Biologists -- on Functions and MR

The task of this section is to examine functional thinking within biology more generally. This serves to reinforce the point of the section following it, that multiple *realizability* (the idea that functions can in principle be realized in different ways) is an important concept within the life sciences, understood as non-reductive enterprises. Functional thinking has many guises in biology. Some (i) are centred in evolutionary approaches; others (ii) in reverse engineering methodologies, where bio systems are compared with mechanical and artificial ones which do similar jobs; others (iii) in robustness analysis, where functions are discovered/hypothesised and scientists investigate how the behaviour is kept stable across perturbing conditions.

i. Evolution and Non-Reductive Causal Explanation

A famous assertion of explanatory pluralism in biology is Ernst Mayr's (1961) distinction between *proximate* and *ultimate* causal explanation. The former refers to mechanisms operating within a living organism, while the latter refers to the evolutionary function or 'purpose' of a behaviour, such as migration. Mayr's larger agenda was to demonstrate that evolutionary biology was a non-replaceable complement to reductive molecular biology that was advancing rapidly at that time. As Beatty (1994:339) writes, "Mayr allowed that the study of proximate causation in biology

[6] Marr (1982: 27) uses a similar analogy: "trying to understand perception by studying only neurons is like trying to understand bird flight by studying only feathers: It just cannot be done. In order to understand bird flight, we have to understand aerodynamics; only then do the structure of feathers and the different shapes of birds' wings make sense."
This passage often read as an assertion of the autonomy of perceptual psychology from neuroscience. I think we should take him to be making same point as Barlow (1961).

approaches ‘the ideal of a purely physical or chemical experiment’ (Mayr 1961, p. 1502). That leaves the evolutionary perspective most responsible for the special character and autonomy of biology.”

Another payoff of a pluralism which accommodates explanations referring to evolutionary or ecological causes is that if demystifies teleology. Apparently goal-directed processes have been selected for because of their adaptive value and so Darwinian evolution offers an account of “purposefulness” being so widespread in the living world.[7] It is convenient to describe subsystems of an organism with recognisable purposes as having “functions”. Important goals of biological research are to construct hypotheses about the functions of structures or processes when they are not immediately obvious, and to develop mechanistic explanations of how systems achieve their functions.

ii. Reverse Engineering Biological Functions

Because biological systems have evolved to perform functions that are often comparable to the actions of man-made devices, one research strategy involves taking a design-stance to biology and seeing if a system can be understood in terms of general engineering principles. In the following passage, neurologist Sir Francis Walshe, contrasts reductive strategies with a reverse engineering approach – one that posits functions in the nervous system and then works back to see how the physical constituents realize them. This approach is advocated at length by neuroscientists Sterling and Laughlin (2015), but I refer to this text from 1961 to reinforce the point that these ideas in neuroscience and biology predate functionalism in the philosophy of mind:

> “The modern student finds it difficult to see the wood for the trees…… He does not always have a synoptic concept of the nervous system in his mind …. If we subject a clock to minute analysis by the methods of physics and chemistry, we shall learn a great deal about its

[7] I’m being simplistic for purposes of illustration. See Jablonka and Lamb (2014) on the complex developments of evolutionary theory post-Darwin.

> constituents, but we shall not discover its operational principles, that is, what makes these constituents function as a clock. Physics and chemistry are not competent to answer questions of this order, which are an engineer's task …. Both modes have their place and limitations ; and they complement one another." Walshe (1961:131)

Section 2 noted the explanatory pluralism and functional thinking within neuroscience; here the point is more general, that functional thinking is prevalent in biological research more generally, and reference to computational functions in neuroscience is just one instance of this. In their advocacy of explanatory pluralism, both Mayr and Walshe display their anti-reductive commitments. Pluralism is recommended because of the limitations of micro-explanations which refer only to physical and chemical causes and therefore miss the larger-scale patterns of organisation that the biologist needs to understand.

iii. Biological Robustness

Biological Robustness has been defined as "the ability of a system to sustain its functionality in the face of perturbation" (Kitano 2004). It is often achieved by multiple different mechanisms capable of performing equivalent functions (*degeneracy*). This raises the question of whether multiple realization is therefore a pervasive feature of biological systems. Boone (forthcoming) argues that because of robustness, MR is common in neural systems. But I suspect that such examples of neurons with varying ratios of ion channel density would not satisfy condition *iv* of Polger and Shapiro's "official recipe" and would be considered instead as normal, within-kind variants. However, the fact that their definition excludes these interesting cases of robustness suggests to me that it misses something interesting about the organisation of living systems. In the living world, stasis is death: the working parts of organisms are constantly rearranging themselves and yet in the midst of this flux they must maintain (approximate) functional stability. The maintenance of functional coherence even though the working parts of the system are in a state of constant upheaval is a problem that biology has had to solve many times, hence the prevalence of many kinds of mechanisms for robustness.[8]

Here is a quotation from a recent discussion of motor learning by neuroscientists taking issue with the "reductionist bias" in neuroscience. They explicitly connect robustness with multiple realizability, in order to make the point that fine-grained descriptions of single circuits are of limited explanatory value.

> "it is now known from careful psychophysical work that many distinct learning algorithms operate together to counter the effects of a perturbation during adaptation, even though phenotypically their summed behavior can look like pure error-based cerebellar learning …. This is a further example of multiple realizability….. [I]f there are many ways to neurally generate the same behavior, then the properties of a single circuit at best are a particular instantiation and do not reveal a general design principle." Krakauer et al. (2017:487)

These authors are stating a case quite recognisable to readers of Fodor (1974) – that multiple realization of a behavioural phenomenon implies that fine-grained neuroscientific description of one instantiation will not take the place of psychological description. However, if we consider the other material discussed above we can see how it suggests ways to go further than the Fodor-Putnam characterisation of MR.

Firstly, it is reasonable to predict that biological functions will be multiply realized (in a non-stringent sense, not satisfying P&S's condition *iv*) because natural selection targets the adaptive "goals" of these functions, and because there will be multiple ways to materially achieve those goals. Furthermore, given the flux of biological hardware, multiple pathways towards the same goal will be a design requirement in order to maintain robust functioning.[9] Because the springboard for this argument is the quasi-teleologial character of living systems, rather than the abstract nature of computation, it diverges from the traditional functionalist idea that multiple

[8] In another paper I discuss this at length (Chirimuuta 2017b).

[9] Canalization in developmental biology is another important example of this. See e.g. Noman et al. (2015); Siegal and Bergman (2002); Jablonka and Lamb (2014:258ff)

realizations of psychological functions are to be expected because psychological states are computational ones, and computation is inherently indifferent to material realization.

Now an obvious reply here is that these cases of MR *in the non-stringent sense* are besides the point. The aim of *The Multiple Realization Book* was to convince us of the need for a more stringent definition, and of the fact that the empirical findings do not deliver cases of MR understood stringently. However, it is here that worries rise to the surface about the "ecological validity" of a set of conditions of MR whose primary example is the corkscrew. We hear from the actual science (not the imaginary science of corkscrews) that functional classifications capture regularities not apparent when description is restricted to a more fine-grained vocabulary (be it physical, chemical or neuroanatomical). Such findings avert us to the limitations of purely reductive research agendas in biology and neuroscience. Such cases have enough philosophical interest that they deserve to be the starting point of an account of MR, not ruled out by fiat because they do not meet the requirements of a definition of MR tried and tested on the science of corkscrews.

4. MR 2.0: Nonreductive Physicalism Made Biological

In this section I sketch some salient features of a new version of MR which is grounded in biology and scientific practice.

i. Multiple Realizability, for Science

While Polger & Shapiro's main target is the empirical claims for multiple realization, they are also dismissive about the idea that certain functions are in principle multiply realizable:

> "Radical Multiple Realization comes into play, in our experience, when advocates of multiple realization try to take seriously the empirical viability of their positions and then

> start to worry about whether the evidence is on their side. Facing challenges about the empirical evidence for actual multiple realization, they retreat to the *possibility* of multiple realization – i.e. to multiple *realizibility.”* (p.52)

The point I would like to make here is that the in principle case for multiple realizability is not only a dialectical retreat for functionalist philosophers – it also has a home in non-reductionist scientific methodology.

I like to think of the functional perspective in science as a handbook for cultivating beneficial ignorance. In the quotation above from Barlow (1961) we saw the connection between (1) the positing of a function, (2) the selection of theory or principles relevant to explaining that function, (3) the design of experiments for collection of data most relevant to the chosen theoretical framework, leading finally to (4) an explanation or description which illuminates how the system works. If you assume that the function initially posited is in principle multiply realizable, you get a useful guide to research. It indicates, for instance, what kinds of data you do not need to collect, and what details can be left out of your models. In Barlow’s example of wings and aerodynamics, many of the material properties of feathers would be irrelevant to the investigation because they just give rise to functionally equivalent variants. (In contrast, such factors would be highly relevant to theories of sexual selection in birds.)

The idea of “canonical neural computations”, promoted by Carandini & Heeger (2012), is an example of this methodology being employed in recent neuroscience. Here the assumption that primary sensory cortex neurons are *linear filters* (the functional posit), and the attendant uptake of signal engineering theory permits scientists to ignore physiological data except for stimulus-response relations for purposes of modelling these neurons, because there are numerous functionally equivalent ways to build a linear filter from neural tissue. In short, positing the multiple realizability of functions allows scientists to black-box details that are irrelevant to their research programmes. Functionalist philosophers of mind overstated case for multiple realizability by claiming that the brain could be made out of any old stuff, but neglected this important methodological point.

ii. Life Gets in the Way

However, it can be hard to disentangle the commitments of this neural computational approach from philosophical functionalism, with its indifference to whether the computation is occurring in artificial or living systems.[10] And if the task of disentangling is neglected, then much of the evidence marshalled by Polger and Shapiro can be presented against neural computationalism of the sort defended by Carandini and Heeger. For in Chapters 5 and 6 of the *MR Book* we learn that operations performed by the nervous system are not in fact indifferent to the details of their realization in the way that functionalist dogma had supposed.

So how is it possible to assert that MR is an important concept for the science of the mind-brain, thus holding on to the non-reductive insights that the concept facilitates, but dropping the empirically untenable commitments of philosophical functionalism? The key move here is to dispense with the indifference to biological hardware which characterises philosophical functionalism, and also much of cognitive science. It can both be true that the material from which the nervous system is built (i.e. living, metabolising cells) is crucial to their function *and* that those functions are multiply realised. My argument here will recapitulate points made above in the previous section.

[10] Here it is tempting to read Carandini through the filter of philosophical functionalism:

> "research in neural computation does not need to rest on an understanding of the underlying biophysics. Some computations, such as thresholding, are closely related to underlying biophysical mechanisms. Others, however, such as divisive normalization, are less likely to map one-to-one onto a biophysical circuit. These computations depend on multiple circuits and mechanisms acting in combination, which may vary from region to region and species to species. In this respect, they resemble a set of instructions in a computer language, which does not map uniquely onto a specific set of transistors or serve uniquely the needs of a specific software application." (Carandini 2012:508)

However, I think we should focus on the methodological point that is being made by means of the computer analogy: that neuroscience will be advanced by application of meso-level descriptions, intermediate between very fine-grained biophysical ones and very coarse-grained behavioural ones (p.507).

First, living material is inherently Heraclitean – it maintains its integrity in the face of thermodynamic forces working against it, through the continual turnover of matter and energy. Unlike the hardware of a computer which is engineered to resist material change when used, neural tissue -- like all living tissue -- keeps changing as it is working. As Godfrey-Smith (2016) argues, it is plausible that this and other material properties of biological brains are key to understanding how cognition and awareness occur in animals, and so radical functionalism is false. At the same time, the Heraclitean nature of the nervous system is a roadblock to the success of purely reductive methodologies. Biological systems robustly maintain their functional profiles in spite of constant internally and externally generated perturbations. Therefore, the functionally relevant patterns -- which stay the same across lower level changes, like the spiral shape of a raging tornado – will not be readily apparent at the finest grain descriptions of individual cells, or even small circuits. Hence the need for "meso-level" descriptions which make salient the shape of the storm against the swirling flux of background changes.[11]

iii. Towards Identity Theory 2.0

While Polger and Shapiro come out in favour of a traditional identity theory,[12] they appear to be bothered by the prospect of a reductionist race to the bottom. Chapters 9 and 10 are intended to head off this threat. Since P&S resist eliminativism regarding mental states and assert the "actual autonomy" of psychology (p.210), it would seem that they need to entertain anti-

[11] Some important criticisms of Sebastian Seung's connectome project to generate synapse-by-synapse reconstructions of retinal circuits centres on this point (Morgan and Lichtman 2013).

[12] "Explanatorily important mental process kinds can be identified with brain process kinds, that is, brain process kinds that can be fully characterized using the resources of the neurosciences." (Polger and Shapiro 2016:26). This brief characterisation of identity theory leaves open questions about what kinds of brain processes, and at what grain of description, psychological kinds are to be identified with – questions which naturally arise from my discussion above. Note also that in the decades since the first identity theory was formulated, neuroscience and psychology have grown towards one another and merged in places. Some neuroscience is a lot like psychology; some psychology is very neuro-sciency. Merely referring to what can be "characterized using the resources of the neurosciences" says very little about how far down the reduction is intended to go.

reductionism of some sort. Obviously, they have rejected MR 1.0 as an anti-reductionist ingredient, but I wonder if they would be tempted by MR 2.0. The view I sketched combines the anti-reductionism of functionalism with the neuro-centrism of identity theory, but it emphasises biology in a way that is atypical of most identity theories.[13] As Godfrey-Smith (2016:fn 32) points out, it is moot whether one considers such a view as a rejection of or a modification of functionalism; and it is equally open to classification as a variety of identity theory.

5. Three Challenges to Polger and Shapiro

In this last section I put larger issues aside and focus on three specific points of tension between P&S's account and explanatory practice in current neuro- and cognitive science.

i. *Computational Explanation without Multiple Realizability?*

In Chapter 8 Polger and Shapiro account for the prevalence of computational explanations in the cognitive sciences. Their claim is that such models are not ontologically committing (p.162), and they insist that any inference from the use of computational models to claims about the putative computational nature of mental states would be to mistake the abstractness of the model for an inherent property of the model's target (p.156).

However, there are reasons to think that P&S *do* have trouble accounting for computational explanations. Let us take as an example vision scientists Frisby and Stone's discussion of lateral inhibition in the eyes of humans and horseshoe crabs. They write that,

> "only a computational theory provides a plausible explanation of why the eyes of both organisms …. should exaggerate luminance edges, and therefore why both organisms should 'suffer' the Chevreul illusion. In essence, we no longer have to accept a mere physiological mechanism as an adequate explanation for a given visual phenomenon, we can now demand that there should also be an underlying computational reason for the

[13] Feigl (1958) is an exception.

> nature of the information processing provided by that particular mechanism." (Frisby and Stone 2012:1050).

On the face of it, this does not sound very non-committal. The suspicion is confirmed if we consider that such explanations work by showing that both of these biological eyes fall into a similarity class which also contains man-made devices performing lateral inhibition, and a completely abstract coding scheme describing this computation.[14] It is important that we be persuaded that there is a genuine similarity between these eyes and between the models – i.e. that the similarity is not just an artefact of the abstractions we've introduced via our models.[15] To extent that computational explanation assumes genuine similarities, then it is ontologically committing. This is not in strong sense of declaring, 'the human retina *is* fundamentally a computing circuit of type L'; rather, in the sense that the abstract, computational description captures (even if in a caricatured or distorted way) something 'real' about the inherent structure or nature of the system.

ii. Not so Beneficial Ignorance

One of P&S's central cases against MR is Karten's neuroanatomical findings of surprising similarity in the circuits specialised for audition in bird and mammal brains. Karten reports that "the avian brain contains cells and circuitry which are nearly identical to those in the mammalian cortex, but disposed as nuclei rather than layers with interlaminar reciprocal connections" (Karten 2013:R15; quoted P&S p.115-116). According to P&S the avian and mammalian brain areas do not count as multiple realizations of the circuit for sensory processing because "the differences

[14] See Chirimuuta (2017a). Cf. Batterman and Rice (2014) on "minimal model explanation".

[15] To make this clear, consider a case in which a similarity or sameness judgment is due to an artefact of abstraction. If I take black and white photographs of oranges and limes, and based on these representations say that the fruits are the same colour, my judgment would be grounded on an artefact of abstraction. My photographic representations reduced the dimensionality of pixel variation from trichromatic to monochromatic, and hence the resulting sameness is artefactual.

in the spatial organization of avian and mammalian brains… are not relevant differences" (P&S p.117).

However, this is a risky inference because it might well be that the spatial organization *is* relevant to sensory processing. One reason to think spatial organisation is relevant stems from the fact that spatial organisation affects axonal wiring length, and this in turn influences the speed and cost of neural information processing (Sterling and Laughlin 2015). So different arrangements might privilege different 'solutions' to problem of auditory processing. Karten's text does not rule out this possibility, though his anatomical drawing (reproduced P&S p.116) does imply that the difference in spatial organisation is irrelevant, because both the mammalian and avian arrangements can be represented by roughly the same circuit diagram.

Polger and Shapiro can of course reply that it is an obvious point that all inferences are risky given the limitations of current knowledge of the brain; we all have to live with risk. Yet the deeper concern is that P&S's "official recipe" exploits the gaps in our knowledge in order to make quick work with potential cases of MR. The condition relevant to Karten's case is (iii), that "factors that lead the As and Bs to be differently classified by S2 [here, neuroanatomy] must be among those that lead them to be commonly classified by S1 [here, computational neuroscience]" (p.67). It is not only neuro-computational models and representations that are abstract. Neuro-anatomical schematic drawings, such as Karsten's abstract from the complexity of actual brain tissue by making assumptions about what structures are the functionally relevant ones.[16] There are plenty of biophysical differences between birds' and mammals' brains that an advocate of MR might point to as potentially relevant to function, but P&S are presuming to be irrelevant, and hence not the ones leading the two kinds of circuits to be commonly classified with respect to sensory processing functions.

[16] It is an obvious point, but still worth stating, that Karten's diagram assumes the neuron doctrine -- that single neurons are the basic units for information processing in the brain of mammals and avians. For this reason no glial cells are depicted. This is a standard assumption but see Cao (2014) for concerns about it.

It is worth returning to the problem of P&S's reliance on toy examples, and why it is significant that brains are not like corkscrews. Even with elaborate Alessi corkscrews it is obvious after a bit of tinkering what the functional parts are, and what is mere aesthetic flourish. With the brain, the separation between information processing structures, and metabolic supports systems is probably more blurred than has commonly been assumed (Cao 2014). This means that even the way that we draw the line between information processing kinds and realizing structures depicted in neuro-anatomy is contestable and theory-dependent. The denier of MR (as defined by the "official recipe") will always have latitude to contest claims of MR because there will never be a neutral way to characterise information processing kinds as opposed to structural and metabolic support, no matter how much data comes in the future.

iii. Deep Networks

Finally, I would like to consider one case that seems to be an obvious instance of multiple realization, but for which I have doubts about the applicability of P&S's conditions for MR. Polger and Shapiro write that:

> "connectionist networks are almost always simulated or emulated using traditional symbolic computing machines – thereby illustrating the fact that their operation is suitably independent of the details of their realization." (p.159)

The idea here is that an artificial connectionist network is the kind of thing that shows independence from its material realization. It would seem to follow straightforwardly that an artificial network for visual object recognition, running on a desktop PC, and the biological networks in the primate ventral stream are multiple realizations of the same functions.

However, visual neuroscientists Yamins and DiCarlo who use many layered connectionist networks ("deep networks") to model the responses of the visual cortex emphasise the "structural" similarities between the artificial and biological structures:

> "Do such top-down goals [e.g. visual object recognition] strongly constrain biological structure? Will performance optimization imposed at the outputs of a network be sufficient to cause hidden layers in the network to behave like real neurons in, for example, V1, V4 or IT? A series of recent results has shown that this might indeed be the case." (Yamins and DiCarlo 2016:359)

In particular, both the artificial and biological networks are hierarchical, many-layered, and with feedforward connections; the individual artificial neurons also develop receptive field properties like their biological counterparts. Yamins and DiCarlo's point is that the task of visual object recognition may be such that any computational network will have to employ these same structural features in order to solve it.

This raises the question of whether the biological and artificial visual networks are multiple realizations after all. P&S (p.159) imply that difference in material hardware is enough to show MR here – and this fits the traditional functionalist intuition that MR just occurs when the same computational function is run on electronic, mechanical or neural hardware. However, when discussing their official recipe, P&S mention as a "litmus test" for MR the question of "whether the same mechanical explanation can explain the operation of the two devices" (p.64). (Answer: in the case of waiter's and double-level corkscrews it cannot, hence there is MR.) It would seem to follow that in the connectionist network case the analogues of "mechanical explanation" are the biophysical and electronic engineering explanations of the hardware of the brain and PC respectively. But, as should be clear from my discussion above,[17] those low level descriptions don't go very far in explaining how the systems work (and this is another problematic point of disanalogy with the corkscrew case). In order to explain object recognition, neuroscientists like Yamins and Carlo refer to the more abstract structural features of the system such as it being hierarchical and many layered. And it turns out that the "structural explanation" of both the artificial and biological networks are the same. So perhaps, by P&S's lights, this turns out not to be a case of MR after all.

[17] And see also Carandini (2012); Jonas and Kording, 2017.

6. Conclusion

The Multiple Realization Book rewards careful consideration and promises to change the landscape of philosophy of mind and cognitive science. While its aim is destructive – to overturn the received wisdom of the discipline – the project is carried out with real creative energy. This energy is infectious and has certainly prompted me to look with excitement at future possibilities for theorising about the mind-brain sciences, some prospects for which I hope to have conveyed in this brief essay.

Acknowledgments

I would like to thank Ken Aizawa for organising the book symposium at the Southern Society for Philosophy and Psychology 2017 where this work was first presented. Credit is also due to my fellow speakers, Trey Boone and Dan Weiskopf, and of course the authors themselves -- Tom Polger and Larry Shapiro.

References

Barlow, H. B. (1961) Possible principles underlying the transformation of sensory messages. In W. A. Rosenblith (Ed.), *Sensory Communication*

Batterman, R. and C. Rice (2014). Minimal model explanations. *Philosophy of Science 81*(3), 349–376.

Beatty, J. (1994). The proximate/ultimate distinction in the multiple careers of Ernst Mayr. *Biology and Philosophy*, *9*, 333–356.

Boone, W. (forthcoming) "Multiple realization and robustness" in M. Bertolaso, S. Caianiello and E. Serrelli (eds.) *Biological Robustness.* Berlin: Springer.

Cao, R. (2014) "Signaling in the Brain: In Search of Functional Units ". *Philosophy of Science* 81(5): 891-901

Carandini, M. (2012). From circuits to behavior: A bridge too far? *Nature*, *15*(4), 507–509.

264, 1333–1336. Carandini, M., & Heeger, D. J. (2012). Normalization as a canonical neural computation. *Nature Reviews Neuroscience*, *13*, 51–62.

Chirimuuta, M. (2017a forthcoming) "The Development and Application of Efficient Coding Explanation in Neuroscience" in Saatsi and Reutlinger (eds.) *Explanation Beyond Causation.* Oxford: Oxford University Press.

Chirimuuta, M. (2017b forthcoming) "Crash Testing an Engineering Framework in Neuroscience: Does the Idea of Robustness Break Down?" *Philosophy of Science*

Chirimuuta, M. (forthcoming) “Vision”. In Sprevak and Colombo (eds.), *The Routledge Handbook of The Computational Mind*

Feigl, H. (1958). “The ‘Mental’ and the ‘Physical’,” in H. Feigl, M. Scriven, and G. Maxwell, (eds.,) *Minnesota Studies in the Philosophy of Science, Volume 2: Concepts, Theories, and the Mind-Body Problem*. Minneapolis: University of Minnesota Press.

Fodor, J. 1974. Special sciences, or the disunity of science as a working hypothesis.
Synthese 28: 97-115.

Frisby & Stone (2012) “Marr: An Appreciation” *Perception.* 41: 1040-1052

Godfrey-Smith, P. (2016) “Mind, Matter, and Metabolism”. *J. Phil.* 113(10):481-506

Jablonka, E. & M. Lamb (2014) *Evolution in Four Dimensions.* Revised edition. Cambridge MA: MIT Press.

Jonas, E., and Kording, K. (2017). Could a neuroscientist understand a micro- processor? PLoS Comput. Biol. *13*, e1005268.

Kaplan, D. M. (2011). Explanation and description in computational neuroscience. *Synthese*, *183*, 339–373.

Karten, H. (2013). “Neocortical Evolution: Neuronal Circuits Arise Independently of Lamination”. *Current Biology* 23(1):R12-15.

Kitano, H. (2004). Biological Robustness. *Nature Reviews Genetics*, 5(11), 826-837

Krakauer, J. W., A. A. Ghazanfar, A. Gomez-Marin, M. A. MacIver, and D. Poeppel (2017) "Neuroscience Needs Behavior: Correcting a Reductionist Bias" *Neuron* 93:480-490

Marr, D. 1982. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. San Francisco: Freeman.

Mayr, E. 1961 "Cause and Effect in Biology" *Science* 134, 1501-1506.

Mitchell, S. D. (2009). *Unsimple truths: Science, complexity, and policy*. Chicago: University of Chicago Press.

Morgan, J. L. and J. W. Lichtman (2013) "Why not connectomics?" *Nature Methods.* 10(6):494.

Noman N, Monjo T, Moscato P, Iba H (2015) Evolving Robust Gene Regulatory Networks. PLoS ONE 10(1): e0116258. doi:10.1371/journal.pone.0116258

Polger, T. W. and L. A. Shapiro (2016). *The Multiple Realization Book*. Oxford: 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.

Siegal and Bergman (2002) canalization paper…

Sterling, P. and S. B. Laughlin (2015). *Principles of Neural Design*. Cambridge, MA: MIT Press.

Walshe, F. (1961) “Contributions of John Hughlings Jackson to Neurology: A Brief Introduction to His Teachings” *Archives of Neurology* 5:119-131

Yamins, D. L. K. and J. J. DiCarlo (2016) “Using goal-driven deep learning models to understand sensory cortex” *Nature Neuroscienc* 19:356–36. doi:10.1038/nn.4244