

Realization, explanation and the mind-body relation

Editor's introduction

Jacqueline A. Sullivan

Received: 15 September 2010 / Revised: 30 September 2010 / Accepted: 1 October 2010 /
Published online: 17 November 2010
© Springer Science+Business Media B.V. 2010

This volume brings together a number of perspectives on the nature of realization explanation and experimentation in the 'special' and biological sciences as well as the related issues of psychoneural reduction and cognitive extension. The first two papers in the volume may be regarded as offering direct responses to the questions: (1) What model of realization is appropriate for understanding the metaphysics of science? and (2) What kind of philosophical work is such a model ultimately supposed to do?

In his contribution to the volume, Carl Gillett responds to the first question by way of an astute critique of the 'subset' model of 'realization'. The subset model, with its origins in philosophy of mind, puts forward two necessary conditions for realization, namely that (i) realized properties and realizing properties have to be present in the same individual and that (ii) the causal powers used to individuate a realized property must be a subset of those causal powers contributed to that property by the realizer. By appeal to an example of a mechanistic explanation of a diamond scratching glass, Gillett demonstrates that when we attempt to use the subset model to capture the 'making-up' relations posited by such explanations, both conditions for realization fail to be satisfied. First, diamonds exhibit the property of hardness and in virtue of this property have the causal power to scratch glass. They have this property and this causal power because they are 'made-up' of carbon atoms that have the causal powers of close bonding and alignment. Yet, carbon atoms are themselves not hard, so condition (i) that realized and realizing properties (i.e., "hardness") must exist in the same individual fails to be satisfied. Furthermore, a diamond may exhibit the causal power to scratch glass, but this causal power is not a subset of the causal powers of

J. A. Sullivan (✉)
Department of Philosophy, University of Alabama at Birmingham, HB 414A, 900 13th Street South,
Birmingham, AL 35294-1260, USA
e-mail: jas1@uab.edu

the carbon atoms that make up that diamond, because individual carbon atoms cannot scratch glass. Thus condition (ii) also fails.

Gillett claims that the primary limitation of the subset model is that it lacks descriptive accuracy with respect to the nature of realization or ‘making-up’ relations in the sciences. For example, it fails to capture that realization relationships are compositional, non-causal, and many-one. However, the problem that he takes to be most fundamental is the failure of the model to capture the “qualitative distinctness” of the properties and causal powers of entities situated at different levels in the ontological hierarchy that enter into realization relationships. Diamonds are hard but the carbon atoms that make them up are not. Atoms can enter into bonding and alignment relationships with other atoms, diamonds cannot. Diamonds can scratch glass, but carbon atoms cannot.

In response to the reader who may wonder why the ‘subset model’ of realization has historically garnered widespread support despite its apparent failings with respect to actual scientific case studies, Gillett points out that proponents of the subset model often conceive of higher-level properties (e.g., “hardness”) as being microbased in structural relationships (e.g., “alignment” and “bonding”) that hold between constituents (e.g., atoms) at a lower level (e.g., COMBO). Thus, they would regard the hardness of a diamond as that structural property of having carbon atoms in certain bonding and alignment relationships. This assumption thus enables them to ascribe two properties (conceived to be situated on different ontological levels) to the same individual.

However, Gillett finds this move by proponents of the subset model unpalatable for several reasons. First, it is difficult to make sense of how structural properties can be said to contribute causal powers, since the powers (e.g., bonding and alignment) are characteristic of the entities that are part of the structure, not of the whole structure itself. Second, the model is “ontologically profligate” in so far as it posits an additional layer of properties in the world and thus two distinct kinds of “making-up” relations (i.e., atoms and their causal powers and the structural property itself), when actual mechanistic explanations in the sciences posit only one layer. Third, it also fails to accommodate the qualitative distinctness of entities, properties, processes and causal powers situated at different ontological levels.

Gillett goes on to show that attempts to use the subset model of realization to block ‘special science’ kinds (e.g., mental states) from ontological reduction also fail. Ontological reductionism is typically understood as the idea that whole entities (i.e., “the realized”) may be reduced to those component parts that make them up (i.e., “the realizers”). As Gillett explains, the subset model of realization implies that such ontological reduction cannot go through because subset-realized properties may be construed as parts of those properties that subset-realize them. He appeals to Shoemaker’s (2001) use of Yablo’s (1992) pigeon Sophie to make the point. Shoemaker claims that when Sophie, who has been trained to peck at red targets, later pecks at a scarlet target, the property of being red (i.e., the “subset realized property”) may be construed as a part of the property of being scarlet (i.e., the realizing property). For, it is the property of the target being red that contributes the relevant causal powers—the powers that cause Sophie to peck at it—rather than the property of being scarlet. However, given that the causal powers used to individuate the realized property (i.e., red) are only a subset of those causal powers contributed by the realizer (i.e., scarlet)

we cannot reduce the former to the latter, because the causal powers of the property of scarlet are greater than those causal powers of the property of being red. Thus, reduction does not go through.

Gillett refers to this as the ‘Overmatching Defense’ and he points out that given an ontological reductionist’s concern with parsimony, there would be no reason to posit higher-level properties like “red” if the causal powers of that property are included in the causal powers of that property that realizes it (i.e., “scarlet”). For, to do so would be, again, in Gillett’s words, “ontologically profligate”—if one of the properties can do all the relevant causal work, there is no need to posit two. Gillett thus concludes that the subset model of realization provides “a more obvious argument for ontological reduction” than even the ontological reductionists themselves provide.

Gillett’s critique of the subset model thus yields at least two important insights. The model has its origins in the wrong place, namely the philosophy of mind, which disables it from being descriptively accurate with respect to those making-up relations that we encounter in mechanistic explanations in the sciences. This suggests that any model that originates outside the metaphysics of science will miss the mark in terms of capturing those relations. The corresponding point is that models of realization that have their origins in the metaphysics of science are to be preferred. Thus, with respect to the issue of what function a model of realization is supposed to serve, Gillett thinks that descriptive accuracy with respect to the “making-up” relations in the sciences is key and is also fundamental for the success of ‘second-order’ projects in the philosophy of mind geared to establish mind-body claims. As he indicates, such descriptive accuracy is to be found in his own account of realization, “the dimensioned view”, which he has developed in other contexts (e.g., [Gillett 2002, 2003](#)).

In his contribution to the volume, Tom Polger responds to the question of what model of realization is appropriate for capturing the metaphysics of science by way of an astute critique of Gillett’s dimensioned view of realization for its failure to *explain* the realization relations at the heart of mechanistic explanations in the sciences. Although Polger shares Gillett’s naturalistic inclination that a proper understanding of the nature of realization must be informed by and accommodate actual scientific case studies, he demonstrates that Gillett’s dimensioned view fails to capture those aspects that set realization apart as a unique kind of dependence relation into which entities situated at different ontological levels may enter. More specifically, he identifies where Gillett’s dimensioned view of realization misses the mark when it comes to understanding the nature of mechanistic explanation in the sciences.

Polger begins his piece by identifying the similarities between philosophical theories of realization and recent models of mechanistic explanation. Realization, according to Polger, is traditionally understood by philosophers of science and mind as an *ontological, non-causal, invariant, asymmetric, existential, constant* and *counterfactual supporting dependence relation* that holds between entities or sets of entities situated at different ontological levels. Philosophers like Gillett, Polger claims, have understood mechanisms (e.g., [Machamer et al. 2000](#)) to qualify as one type of entity that can factor into this kind of dependence relation. Polger refers us to Gillett’s appeal to Craver’s (e.g., [2002](#)) example of the mechanism of spatial memory. On this example, the mechanism in the hippocampus (i.e., “hippocampal mechanism”) by which a spatial map is formed is brought about by the “synaptic mechanism” of

long-term potentiation, which is brought about by the “mechanism of ion channels opening” in those cells that comprise the synaptic membrane. Polger agrees that it is tempting, in light of Machamer et al.’s account of mechanisms, to understand these relationships between lower-level mechanisms and higher-level mechanisms as one of *realization*. He claims that it is precisely these kinds of “vertical dependence relations” between mechanisms that Gillett and Aizawa have sought to capture by virtue of Gillett’s dimensioned view (Gillett 2003, 2007; Aizawa and Gillett 2009).

Polger points out that identifying the ways in which traditional accounts of realization and mechanisms are similar may be an interesting descriptive project for the metaphysics of science if it were not for the fact that new mechanists like Machamer et al. actually *presuppose* that the entities and activities that comprise mechanisms stand in dependence relations to one another. After all, scientists themselves have to presuppose these kinds of dependence relations, in order for mechanistic explanations to even get off the ground. This must mean, Polger intimates, that there is more to mechanistic explanation than the “vertical inter-mechanism dependence relations” that Gillett’s dimensioned view of realization descriptively accommodates.

In order to identify what that something more is, Polger introduces a novel distinction between descriptive and explanatory approaches to realization. On this distinction, a *descriptive approach* to realization understands it as “the only metaphysically basic non-causal ontological dependence relation among properties at different mereological levels” (Polger, this volume). Polger identifies Gillett’s dimensioned view as the most prominent example of this type of approach in so far as it posits that only properties situated at different ontological levels can stand in realization relations to one another. Thus, the property of close bonding in carbon atoms on Gillett’s view may be understood as realizing the property of hardness in the diamond. In contrast, on an explanatory approach, realization “is only one among several ontological dependence relations that properties might stand in relation to one another” (Polger, this volume). What is unique about explanatory approaches to realization is that they treat the dependence relationship as something distinct from mere ontological dependence. Polger emphasizes the importance of drawing this distinction because it accommodates the fact that some of the properties that we ascribe to entities are exclusively *functional* properties whereas some are not. To take Polger’s example, a lawn mower has the property of being able to cut grass (i.e., a functional property) as well as the property of having a mass (i.e., a non-functional property). Yet, how the former property is realized differs from how the latter property is realized, and any viable account of realization ought to discriminate the difference.

After turning back a potential set of objections to the descriptive-explanatory realization distinction, Polger goes on to demonstrate, by appeal to an example, why explanatory approaches to realization in the metaphysics of science are to be preferred when it comes to understanding the nature of mechanistic explanations. Polger reminds the reader that mechanists like Machamer et al. distinguish mechanisms from mere aggregates in so far as they define mechanisms “as entities and activities organized such that they are productive of regular changes from start or set-up to finish or termination conditions” (Machamer et al. 2000, p. 3). Referring back to the example of the lawn mower, Polger points out that a disassembled set of lawn mower parts and a fully assembled lawn mower may both be considered as having the same overall

“aggregate properties”. For example, the pile and the assembled lawn mower may be said to have the same mass and to take up the same amount of space. However, what distinguishes the assembled lawn mower from the pile of parts, Polger tells us, is that it has properties over and above those properties that the pile and the assembled lawn mower share in common. Specifically, its parts are organized in such a way that they may bring about the functional property of cutting grass, whereas the pile of lawn mower parts lack this organization and as a consequence lack the functional property as well. Yet, Polger claims, it is just such functional properties that mechanistic explanations in the sciences are more often than not put forward to explain. Gillett’s dimensioned view of realization, in so far as it is “indiscriminate” with respect to how different properties of entities come about, cannot be used to capture this fundamental difference between mechanisms and mere aggregates. This is ironic, Polger tells us, because Gillett touts the dimensioned view of realization as the most descriptively accurate one for capturing the making-up relations at the heart of mechanistic explanations, yet as it turns out, the model is inappropriate for this purpose. However, not only is the model descriptively inaccurate, but also it cannot be used for what Polger takes to be the more fundamental aim of philosophical models of realization, namely to explain why two things may be said to stand in a realization relationship to one another as opposed to standing in some other kind of relationship.

Thus, in response to the question of what model of realization is to be preferred to capture the metaphysics of science, Polger puts forward the explanatory approach. This approach is superior because it can be used to explain those realization relations at the heart of mechanistic explanations in the sciences. It is no real philosophical feat, Polger tells us, for a model of realization to merely describe those ontological dependence relationships on which mechanistic explanations depend. Rather, to discover that two things stand in a realization relationship to one another ought to be, in Polger’s words, “a philosophical achievement in the same way that providing a mechanistic explanation is a scientific achievement”.

Philippe Huneman’s contribution to the volume may be regarded as a direct response to debates about the nature of realization in the metaphysics of science as instantiated in Gillett’s and Polger’s articles. To date such debates have exclusively sought to understand the realization relationships at the heart of mechanistic explanations. Yet, Huneman tells us, there are other important types of explanation in the sciences that models of realization ought to be able to accommodate and this raises the possibility that no single model of realization will be able to capture all of them. One such important type of explanation that is the primary focus of Huneman’s paper is what he designates “topological explanation”.

While Huneman does not deny that mechanistic explanations are ubiquitous throughout the biological and social sciences, he appeals to case studies from ecology and evolutionary biology in order to demonstrate that topological explanations are equally as pervasive. Huneman begins by suggesting that a basic feature that sets topological explanations apart from mechanistic explanations is that the explanandum phenomenon is always a target property, trait or outcome of a system, rather than an activity (i.e., a functional property). A second feature that may be used to differentiate them is that the explanans of a topological explanation involves an exclusive appeal to the “topological properties”—certain traits, properties, behaviors or

features—exhibited by that system. These properties are typically not concrete; rather, they are represented abstractly in network models or graphs, which are then used to explain the target property of interest.

Huneman then presents a series of examples of explanations from ecology and evolutionary biology that instantiate these basic features. One example involves the explanation of topological properties of an ecological community. An ecological community, according to Huneman, may be understood as a system comprised of a number of different species of organism. These species prey on each other, and the predation relationships between and among the groups that comprise the community emerge as topological properties of the entire community. These topological properties are not directly observable features of the community yet they may be represented and understood within an abstract mathematical space, such as a network or a graph. Huneman explains that predation relationships may differ from one ecological community to the next, and in instances in which they do differ the evolution of frequencies of the species that comprise a given community will be impacted. When an investigator aims to explain this impact, Huneman asserts that an explanation that describes the causal interactions between the individual parts (i.e., the species) that comprise an ecological community will be insufficient. What is required instead is an explanation that appeals to the topological properties of the system, namely, the predation relationships among the species. In such an explanation, the presence of some feature of a system (i.e. evolution of frequency) is explained by reference to the topological properties of that system (i.e., predation relationships). Thus, the explanation is topological rather than mechanistic.

Huneman then appeals to [Machamer et al. \(2000\)](#) account of mechanistic explanation in order to further illuminate what sets topological explanations apart. Whereas mechanistic explanations include descriptions of activities in the explanans and often in the explanandum, topological explanations only aim to explain properties of a system by reference to other properties of that same system. Second, topological explanations do not involve an appeal to “a chronology of events” from “starting conditions to termination conditions” in order to explain a target property ([Machamer et al. 2000](#), p. 3). The target property is simply explained by the current topological properties of the system. Third, the properties in mechanistic explanations are causally individuated in so far as they are the outcomes of causal processes; the properties in topological explanations are not. Fourth, Huneman suggests that mechanistic explanations are compatible with two kinds of realization that Gillett identifies, M-type realization, which is “an asymmetric determination relation between causally individuated properties” (and the type of realization captured by Gillett’s dimensioned view), and A-type realization, which is simply “an isomorphism taken to hold between properties” (Huneman, this volume). However, according to Huneman, the kinds of realization relationships found in topological explanations are exclusively of the A-type.

Huneman goes on to point out that mechanistic and topological explanations may interact in a number of different ways in the biological and social sciences. First, mechanistic explanations are sometimes prerequisite for providing topological explanations. For example, sometimes identifying causal interactions between species that comprise an ecological community may be relevant for providing a topological explanation of a property of that community (e.g., the stability of that community). In other

instances, both mechanistic and topological explanations may be put forward to explain a non-functional property. Still, Huneman insists that there are some cases in which a topological explanation of a property will be considered sufficient and complete, and the need to provide a mechanistic explanation will be unnecessary. Additionally, he suggests that topological explanations and mechanistic explanations may be mutually constraining. A topological explanation, for example, may identify some topological properties of a system as explanatorily relevant, which will put constraints on what mechanisms will factor into a mechanistic explanation of a property of that system.

Huneman concludes his piece by using the mechanistic-topological explanation distinction to draw conclusions about which model of realization is appropriate for capturing the metaphysics of science. In demonstrating the prevalence of topological explanations in addition to mechanistic explanations in the sciences, Huneman makes a convincing case that at least two models of realization are required. Thus, Huneman may be regarded as arguing that what Polger classifies as descriptive and explanatory approaches to realization are both equally necessary for capturing the metaphysics of science. The difference between Huneman and Polger, however is that Huneman assumes that Gillett's dimensioned view of realization is sufficient to capture the nature of mechanistic explanation in the sciences, whereas Polger demonstrates that it is not. In fact, by the end of Polger's article and in light of Huneman's article, Gillett's dimensioned view appears more amenable for capturing topological explanations of non-functional properties than mechanistic explanations of functional ones!

In his contribution to the volume, John Bickle expresses his dismay with recent debates about realization playing out in the metaphysics of science. He claims that philosophers should not concern themselves with trying to understand ontological dependence relationships that hold between entities (i.e., functional or non-functional properties); rather, they should aim at understanding the relationships between theories and experimental practices of different areas of science. If we want to learn the answer to the question of how the mind is related to the brain, Bickle tells us, we need first and foremost to understand how the sciences of the mind are related to the sciences of the brain.

The primary aim of Bickle's article is to address the interesting question of whether the various waves of challenges during the past 40 years to the argument from multiple realization, which include recent debates about realization in the metaphysics of science, have helped or hindered the aim of reducing psychology to neuroscience and mind to brain. Bickle begins by presenting the traditional argument from multiple realization, which posits the assumption that psychological kinds are multiply realized in physical kinds and which is widely taken to establish the conclusions that the identity theory is false and reductive physicalism is not possible. He then considers two historical attempts to refute the argument that he organizes according to which premise of the argument comes under attack.

Attributing the first wave to Lewis (1967), Bickle claims that Lewis's target was the premise that multiple realization was incompatible with identity theory. In so far as Lewis proposed that identity claims could be relativized to context, this raised the possibility that species-specific identities could be established between psychological and biological kinds. Thus, Lewis took identity theory and multiple realization to be compatible and as Bickle points out, many philosophers of science conceded the

point. However, proponents of multiple realization put forward examples of recovery of function and neural plasticity as evidential support for the idea that the same psychological kind could be realized differently in the same physical system over time. Lewis's species-specific identities could not be used to salvage identity theory in the face of this type of multiple realization, which was used to reassert the stronger premise that psychological kinds are massively multiple realized.

Bickle argues, however, that such arguments achieved success primarily due to a failure on the part of proponents of multiple realization to recognize that all available models of intertheoretic reduction, including that developed by Nagel (1961) were actually compatible with the multiple realization of reduced in reducing kinds both across species and within the same individual. Bickle points out that Nagel, in formulating his connectability condition for intertheoretic reduction actually put forward one-way conditional bridge laws as a viable means to effect reductions between reduced and reducing kinds. Furthermore, proponents of reduction, including Bickle himself, have appealed to cases from the physical sciences in order to show that even though a macro-level property like the temperature of a gas may be subject to different micro-level instantiations over time with respect to the mean kinetic energy of the molecules that comprise it, intertheoretic reduction still goes through. Bickle claims that if even in the context of the most fundamental sciences multiple realization is compatible with intertheoretic reduction, it should certainly not be regarded as an obstacle to the reduction of psychology to neuroscience.

Bickle entertains the possibility that the reason proponents of multiple realization did not recognize that successful reductions could be accomplished in science even in the face of multiple realization had to do in part with Jaegwon Kim's widespread influence in the philosophy of mind. According to Bickle, Kim (1993) accepted that psychological kinds were multiply realizable in physical kinds, but he took multiple realization to be compatible with "structure-specific" reductions that were similar to but narrower than Lewis's species-specific reductions. Bickle suggests that because neither Kim's argument against multiple realization nor his model of structure-specific reduction were informed by actual scientific case studies, proponents of multiple realization took Kim's model to be compatible with the idea that psychology lacked the kind of generality that had been traditionally ascribed to it. Bickle claims that because proponents of MR were not willing to give up the generality of psychology so easily, they rejected Kim's argument for reduction. An additional factor that contributed to the non-reductive physicalism consensus, according to Bickle, was that Kim's argument for structure-specific reductions and against multiple realization came to be regarded by the philosophical community as intimately related to his argument for explanatory exclusion. Thus, philosophers who directed their criticisms at the latter argument regarded themselves as refuting both arguments simultaneously, which is why the philosophical community remained convinced of the incompatibility of multiple realization with intertheoretic reduction.

Whereas the aim of the first wave was to establish the compatibility of multiple realization with reduction, Bickle takes the goal of the second wave to be a refutation of the premise that psychological kinds are multiply realized in physical kinds. Bickle identifies Shapiro (2000, 2003) as the primary instigator of the second wave. Shapiro puts forward a model of realization that ultimately makes the requirements

for establishing multiple realization far more stringent. On Shapiro's model, establishing that a functional kind is multiply realized requires identifying at least two token instances of that kind and demonstrating that the properties by which the two realize that function are different. However, Bickle, appealing to Shapiro's example of the functional kind corkscrew, suggests that if two token instances of corkscrew do not differ with respect to those properties that are causally relevant for bringing the function about, then they cannot be said to be different realizations of the same kind. Thus two waiter's corkscrews, one made of aluminum, the other of steel, would not qualify as multiple realizations of the functional kind corkscrew. On the other hand, Bickle points out that if the two tokens differ with respect to those properties by which the function is brought about, for example, a waiter's corkscrew and a winged corkscrew, then they actually qualify as distinct functional kinds and thus should not be grouped together as instances of the very same functional kind. Bickle interprets Shapiro, then, as providing a convincing argument against the multiple realization of functional kinds.

Despite the success Bickle attributes to Shapiro with respect to undermining the multiple realization of psychological kinds in physical kinds, Bickle claims that because the aim of such arguments is not to directly establish psychoneural reduction, their results are compatible with alternative mind-brain theories, including identity theory (e.g., Polger 2004) and mechanistic reduction (e.g., Bechtel 2009). Furthermore, the second wave has prompted a recasting of the nature of the realization relationship that links the multiple realization debate to debates about scientific realism and the nature of the phenomena posited by the mind-brain sciences. However, Bickle is quick to point out that psychoneural reduction is a thesis about the kinds of relationships that can hold between theories, experimental practices and results of different sciences rather than the kinds of ontological relationships that hold between the entities studied by those sciences. Thus, the two waves of attacks on the argument from multiple realization have been irrelevant to the project of the psychoneural reductionist. However, Bickle takes the success of that project to already be guaranteed. In his own work (Bickle 2003, 2006), he has appealed to real scientific case studies from the field of molecular and cellular cognition in order to demonstrate that psychological functions like memory consolidation can be reduced to the same cellular and molecular mechanisms within individuals and across species (PKA-cAMP-CREB). So, the fact that the waves of attacks on the argument from multiple realization have brought no solace to the psychoneural reductionist does not really matter, because cellular and molecular neuroscience, by Bickle's own lights, has provided him with all the comfort he needs.

My contribution to this volume is prompted by a concern that those case studies used to illustrate mechanistic explanations that have factored into recent debates about realization in the metaphysics of science have as their targets behavioral effects rather than cognitive functions. This would be problematic because it may lead to the mistaken conclusion that contemporary neuroscience is providing reductive explanations for cognitive functions when it is actually the case that these functions have been *eliminated* for behavioral effects. The primary case study that has been imported into recent discussions about the metaphysics of science is the mechanistic explanation of rodent spatial memory in the Morris water maze (e.g., Craver and Darden 2001; Craver 2007). While behavioral effects such as moving towards a target and cognitive

functions such as remembering the spatial location of a hidden target (i.e., “spatial memory”) both qualify as activities in which a rodent in a maze may engage, Craver and Darden’s (2001) and Craver (2007) treatment of the case does little to clarify what the target of the mechanistic explanation is. In fact, despite the consensus in recent philosophy of neuroscience with respect to the importance of experimental methods for delineating explanatory targets (e.g., Bechtel 2008; Bickle 2003, 2006; Craver 2007; Feest Forthcoming; Sullivan 2009), experimental devices like the Morris water maze have not been evaluated with respect to their success or failure in isolating such targets. More specifically, no one has stopped to consider precisely ‘what’ devices like the water maze may be used to delineate. Engaging in this task is the focus of my paper.

I begin by describing a conceptual framework that is intended to facilitate analyses of experiments in cognitive neurobiology like those run in conjunction with the Morris water maze. Given that the water maze is one of the most widely used apparatus for studying “spatial learning and memory”, and the basic abstract structure of the experimental learning paradigm has not changed significantly since its original introduction into the experimental literature, I take it as a legitimate investigative strategy to analyze Richard Morris’s first published research study using the water maze (Morris 1981). Beginning with the original water maze as my target allows me to momentarily avoid the fact that historically the water maze has been attributed with delineating a whole host of cognitive functions and behavioral effects ranging from “spatial memory” to “spatial navigation” to “water maze performance”, to name only a few! This strategy also allows me to speculate as to the reasons why the water maze has, in its ~30 year history, come to be associated with so many different kinds of phenomena.

In my analysis of the case study, I demonstrate that Morris sought to develop an experimental learning paradigm that would successfully dissociate between two types of orientation and two types of learning that were associated with them namely, “distal orientation and place learning” and “response orientation and cue learning” (See for example, Morris 1981; Restle 1957). What is referred to as “the hidden condition” of the water maze was supposed to delineate distal orientation and place learning, whereas “the cued condition” of the water maze was supposed to delineate response orientation and cue learning. I demonstrate that although Morris emerged from his first research study with an experimental learning paradigm that could be used to reliably produce a set of behavioral effects, the paradigm lacked the requisite reliability to discriminate how rats trained in the water maze “oriented to” and “learned the location of” the maze targets (i.e., platforms). Thus, Morris did not successfully dissociate these two kinds of orientation/learning, even though the reliable production and detection of the behavioral effects in the maze was a true scientific achievement. The conclusion of my analysis of the water maze, then, is that it discriminates a set of behavioral effects rather than two discrete cognitive functions.

I go on suggest a possible source of the lack of reliability that I attribute to Morris’s experiments using the water maze. On the one hand we may simply say that Morris’s experiment lacked the requisite experimental controls for reliably discriminating a discrete cognitive function (or two such functions). However, I point to another source, namely an acknowledgement in the literature in experimental psychology pre-dating the water maze by nearly 20 years, which suggested that because rodents will always

use whatever means available to locate a goal object no matter what the context, the concepts of “response orientation”, “cue learning”, “place learning” and “distal orientation” could not be used to capture discrete separable phenomena (Restle 1957). This suggests that the concepts were inappropriate for grounding the construction of an experimental learning paradigm that was intended for the very purpose of circumscribing them (i.e., the water maze). Indeed, when we look at recent work on “spatial navigation” (See for example, Sutherland and Hamilton 2004), this claim still appears to ring true: rats continue to be understood as using a variety of simultaneous strategies to do whatever it is that they do in the water maze. They also are described as learning a whole host of different kinds of things (e.g., Morris and Frey 1997). I appeal to my conclusion that Morris’s original study demonstrated only that the water maze could be used to circumscribe a set of behavioral effects in order to explain why it is the case that across the experimental and review literature since 1981 investigators have put forward numerous different concepts as a means to classify those effects. It would seem that everyone encounters some difficulty in trying to identify precisely ‘what’ the water maze circumscribes.

I conclude the paper by identifying some of the implications of my analysis for recent work on the nature of realization and the mind-body relation in the metaphysics of science. I suggest that a failure to evaluate the methods by virtue of which cognitive functions are purportedly delineated may prompt the incorrect conclusion that mechanistic explanations are being provided for cognitive functions rather than behavioral effects. If mechanistic explanations are understood to be a form of reductive explanation, this may lead a proponent of reductive physicalism to wrongly assume that case studies like that of spatial memory and the Morris water maze may be used as evidential support for that position. However, such case studies may only be used to establish that cognitive neurobiology endorses a form of eliminative materialism that some have referred to as ‘the new behaviorism’ (e.g., Machamer 2009; Sullivan et al. 2004). The moral, then, is that analyses of the experimental methods used in contemporary neuroscience are fundamental for “first-order projects” in the metaphysics of science, such as understanding ‘what’ is being realized in the context of neuroscientific experiments and “second-order projects”, such as understanding what light evidence from neuroscience may be used to shed on the relationships between mind and brain and psychology and neuroscience.

Sven Walter’s contribution to the volume addresses a topic that has also factored into debates about realization in philosophy of mind and the metaphysics of science, namely, *cognitive extension*. The basic idea of cognitive extension is that “the material vehicles of some cognitive processes are spatially distributed over the brain and the extracranial parts of the body and the world” (Walter, this volume). Walter explains that arguments for cognitive extension depend on the truth of the parity principle, which proponents of cognitive extension simply assume. The parity principle in essence asserts that just so long as cognitive processes that go on in the head may be regarded as functionally equivalent to processes that go on outside of it, the external processes may be considered cognitive, too. As Walter illustrates by appeal to Clark and Chalmers’ (1998) thought experiment, according to the parity principle, if two individuals both want to attend an art exhibition at the MoMA and the first remembers where the museum is and relies on her memory to get there, whereas the second has

Alzheimer's and has to rely on a set of directions that he jotted into a notebook before he lost his memory, the proponent of cognitive extension who endorses the parity principle will, given that they take both processes to realize the same function, treat both as possessing the mark of the cognitive (even though only one of them goes on in the head).

After explaining the nature of the parity argument, Walter turns back a series of criticisms of the parity principle. First, he considers the claim that it is schizophrenic of the proponent of cognitive extension, who already assumes that the boundaries of cognition are outside the head, to rely on the parity principle to provide a definition of what it means for a process to be cognitive, precisely because the parity principle appeals to cognitive processes "in the head" as the point of comparison between internal and external processes (Di Paolo 2009). In response to this claim, Walter argues convincingly that the proponent of cognitive extension does not rely on the parity principle as anything like a "real definition of what it means for a process to be cognitive". Rather it is appealed to as a heuristic for determining the functional equivalence between two processes, one that is "uncontroversially" cognitive and the other whose cognitive status is in question. Walter thinks this is a legitimate move, rather than a "schizophrenic" one.

In response to a second criticism (Rupert 2009) that the parity principle cannot be used to determine whether or not two processes are functionally equivalent, Walter claims that this criticism misinterprets the aims of the proponent of cognitive extension's appeal to the parity principle. Walter understands Rupert to interpret the parity principle as asserting something about the physical equivalence between two processes, one that takes place inside and the other outside the head. Yet, Walter points out that the aim of the parity principle is exclusively to say something about what it would mean for the two processes to be functionally equivalent, which has to do only with whether or not they have the same *computational* structure.

Walter considers a final argument against the parity principle that appeals to the dissimilarity between cognitive and extended processes. Whereas cognitive processes in the head are "informationally integrated" with and updated in conjunction with other beliefs, processes that go on outside the head lacks such integration (Weiskopf 2008). However, Walter argues that this argument would only be successful if it were a requirement that cognitive processes be individuated so finely as to make the comparison between processes in the head and processes in the world impossible. However, Walter asserts that there is no reason why the proponent of cognitive extension should not be allowed to select whatever grain of analysis he takes to be requisite for determining the functional equivalence of two processes. So, he takes this criticism of the parity principle to fall short as well.

Having demonstrated that the available criticisms of the parity principle in the philosophical literature fail, Walter offers two novel criticisms of it. The first criticism has to do with the fact that the principle requires cognitive processes to be functionally individuated; after all, Walter points out, this is prerequisite for determining functional equivalence. However, Walter demonstrates that the parity principle by itself does not provide us with criteria for determining functional equivalence between intracranial processes and extended ones. What is supposed to provide us with the requisite criteria, he claims, is functionalism itself in so far as it contains implicit methods for

the identification of cognitive processes and a commitment with respect to “the correct level of grain at which functional equivalence has to be assessed” (Walter, this volume).

Walter’s second criticism has to do with the fact that the parity argument on its own provides us with no “mark of the cognitive” and thus, when we try to reformulate the parity principle so as to get clear on why internal and extended processes should be treated as functionally equivalent, we come to realize that we lack principled grounds upon which to establish whether a process is cognitive or not. Walter offers three alternative formulations of the parity principle in order to demonstrate that where a process takes place (i.e., “in the head”) seems to be the only feature that grounds whether that process or an equivalent process in the world is cognitive or not. However, as Walter convincingly argues via an evaluation of an argument by Sprevak (2010), merely pointing to the location in which a process takes place cannot be used as a basis for calling that process cognitive and further asserting that a process taken to be functionally equivalent is also cognitive. As Walter indicates, many processes go on inside the head that cognitive scientists would not be inclined to designate as cognitive. So there must be something that it means to be a cognitive process that is not captured by either the parity principle or functionalism. Proponents of cognitive extension have failed to identify such a mark. Yet without such a mark, Walter concludes, the project of establishing cognitive extension is doomed to failure.

It is relevant in concluding this introduction to say something about the generating circumstances of the articles contained in this volume. The authors first presented these papers as part of a workshop entitled “Realization, Multiplicity and Experimentation in Biology, Psychology & Neuroscience”, which was held in the Philosophy Department at the University of Alabama at Birmingham (UAB) in February 2010, thanks to Harold Kincaid and the generous support of UAB’s Department of Philosophy and the *Center for Ethics and Values in the Sciences*. The workshop was the latest in a series of what have come to be annual informal workshops aimed at discussing core issues in the metaphysics of science. These workshops have their origin in a workshop organized by Larry Shapiro and held in the Philosophy Department at the University of Wisconsin–Madison in 2006. The composition of the group fluctuates from year to year. Participants in the current workshop included myself, Ken Aizawa, John Bickle, Carl Gillett, Philippe Huneman, Tom Polger, Larry Shapiro, and Sven Walter.

References

- Aizawa, K., & Gillett, C. (2009). The (multiple) realization of psychological and other properties in the sciences. *Mind and Language*, 24, 181–208.
- Bechtel, W. (2008). *Mental mechanisms: Philosophical perspectives on cognitive neuroscience*. New York: Lawrence Erlbaum.
- Bechtel, W. (2009). Molecules, systems, and behavior: Another view of memory consolidation. In J. Bickle (Ed.), *Oxford handbook of philosophy and neuroscience* (pp. 13–40). New York: Oxford University Press.
- Bickle, J. (2003). *Philosophy and neuroscience: A ruthlessly reductive account*. Dordrecht: Springer.
- Bickle, J. (2006). Reducing mind to molecular pathways: Explicating the reductionism implicit in current cellular and molecular neuroscience. *Synthese*, 151, 411–434.
- Clark, A., & Chalmers, D. (1998). The extended mind. *Analysis*, 58, 7–19.

- Craver, C. (2002). Interlevel experiments and multilevel mechanisms in the neuroscience of memory. *Philosophy of Science*, 69(3 Supplement, 2), S83–S97.
- Craver, C. (2007). *Explaining the brain*. New York: Oxford University Press.
- Craver, C., & Darden, L. (2001). Discovering mechanisms in neurobiology: The case of spatial memory. In P. K. Machamer, R. Grush, & P. McLaughlin (Eds.), *Theory and method in the neurosciences* (pp. 112–136). Pittsburgh: University of Pittsburgh Press.
- Di Paolo, E. (2009). Extended life. *Topoi*, 28, 9–21.
- Feest, U. (Forthcoming). What exactly is stabilized when phenomena are stabilized? *Synthese*.
- Gillett, C. (2002). The dimensions of realization: A critique of the standard view. *Analysis*, 64(4), 316–323.
- Gillett, C. (2003). The metaphysics of realization, multiple realizability, and the special sciences. *The Journal of Philosophy*, 100(11), 591–603.
- Gillett, C. (2007). Understanding the new reductionism: The metaphysics of science and compositional reduction. *The Journal of Philosophy*, CIV, 4, 193–216.
- Kim, J. (1993). *Supervenience and the mind*. Cambridge: Cambridge University Press.
- Lewis, D. (1967). Review of art, mind, and religion. *Journal of Philosophy*, 66, 23–35.
- Machamer, P. K. (2009). Neuroscience, learning and the return to behaviorism. In J. Bickle (Ed.), *The Oxford Handbook of Philosophy and Neuroscience* (pp. 166–178). Oxford: Oxford University Press.
- Machamer, P., Darden, L., & Craver, C. (2000). Thinking about mechanisms. *Philosophy of Science*, 67, 1–25.
- Morris, R. G. M. (1981). Spatial localization does not require the presence of local cues. *Learning and Motivation*, 12, 239–260.
- Morris, R., & Frey, U. (1997). Hippocampal synaptic plasticity: Role in spatial learning or the automatic recording of attended experience? *Philosophical Transactions of the Royal Society of London B*, 352, 1489–1503.
- Nagel, E. (1961). *The structure of science*. New York: Harcourt, Brace and World.
- Polger, T. (2004). *Natural minds*. Cambridge MA: MIT Press.
- Restle, F. (1957). Discrimination of cues in mazes: A resolution to the “place-vs.-response” question. *The Psychological Review*, 64(4), 217–228.
- Rupert, R. (2009). *Cognitive systems and the extended mind*. Oxford: Oxford University Press.
- Shapiro, L. (2000). Multiple realizations. *Journal of Philosophy*, XCVII, 635(654), 635–654.
- Shapiro, L. (2003). *The mind incarnate*. Cambridge: MIT Press.
- Shoemaker, S. (2001). Realization and mental causation. In C. Gillett & B. Loewer (Eds.), *Physicalism and its discontents*. Cambridge: Cambridge University Press.
- Sprevak, M. (2010). Functionalism and extended cognition. *Journal of Philosophy*, 106, 503–527.
- Sullivan, J. A., Machamer, P. K., & Thiels, E. (2004). The study of learning and memory then and now: Evidence for conceptual change? Paper presented at the *Society for Neuroscience Annual Meeting*, San Diego, CA.
- Sullivan, J. A. (2009). The Multiplicity of experimental protocols: A challenge to reductionist and non-reductionist models of the unity of science. *Synthese*, 167, 511–539.
- Sutherland, R. J., & Hamilton, D. A. (2004). Rodent spatial navigation: At the crossroads of cognition and movement. *Neuroscience and Biobehavioral Reviews*, 28, 687–697.
- Weiskopf, D. (2008). Patrolling the mind’s boundaries. *Erkenntnis*, 68, 265–276.
- Yablo, S. (1992). Mental causation. *The Philosophical Review*, 101, 245–280.