

Crosscutting psycho-neural taxonomies: the case of episodic memory

Muhammad Ali Khalidi*

Department of Philosophy and Cognitive Science Program, York University, Toronto, Canada

(Received 17 March 2017; final version received 17 March 2017)

I will begin by proposing a taxonomy of taxonomic positions regarding the mind–brain: localism, globalism, revisionism, and contextualism, and will go on to focus on the last position. Although some versions of contextualism have been defended by various researchers, they largely limit themselves to a version of neural contextualism: different brain regions perform different functions in different neural contexts. I will defend what I call “environmental-etiological contextualism,” according to which the psychological functions carried out by various neural regions can only be identified and individuated against an environmental context or with reference to a causal history. While this idea may seem innocuous enough, it has important implications for a structure-to-function mapping in the mind and brain sciences. It entails that the same neural structures can subserve different psychological functions in different contexts, leading to crosscutting psycho-neural mappings. I will try to illustrate how this can occur with reference to recent research on episodic memory.

Keywords: reductionism; taxonomy; cognitive ontology; structure-to-function mapping

Philosophers, as well as neuroscientists, psychologists, and cognitive scientists have recently been preoccupied with the question of psycho-neural taxonomy. Though this debate is related to the traditional metaphysical debate in the philosophy of mind concerning reductionism, eliminativism, functionalism, and other ontological positions, it is also partly orthogonal to it. The more recent debate has to do with the precise ways in which neural categories will map onto cognitive and psychological ones (if at all), what cognitive scientists and neuroscientists often refer to as the “structure-to-function mapping.” This question is *partly* orthogonal to the ontological question because some of the traditional positions concerning mental or cognitive ontology are (despite appearances) compatible with more than one position in this more recent debate. However, it is not entirely orthogonal, since, as I will try to show, at least one position on the taxonomic issue precludes a familiar type of psycho-neural reductionism in ontology.

A number of positions have emerged in the structure-to-function debate, the main ones being: localism, globalism, revisionism, and contextualism (these labels are not widely used but I will try to characterize them in some detail in what follows). In this paper, I will try to make an argument for the last of these positions, contextualism, based on the case study of episodic memory. Instead of the relatively common neural contextualism, I will argue for an environmental-etiological contextualism, which has not been adequately explored. Moreover, I will try to show that contextualism (at least the variety that I will be defending) is

*Email: khalidi@yorku.ca

not compatible with a familiar kind of psycho-neural reductionism, and if vindicated, would undermine it. In Section 1, I will articulate four different positions on taxonomy, attempting to contrast them, with particular emphasis on contextualism. Then, in Section 2, I will try to give a general account of the structure-to-function mapping, based on a particular understanding of the notions of structure and function, and will argue that structural and functional taxonomies can crosscut one another. Armed with this understanding of the structure-to-function mapping, in Section 3, I will examine the case of episodic memory, arguing that it may lend itself to crosscutting neural and psychological taxonomies. Finally, in Section 4, I will conclude that at the current historical juncture different taxonomic practices seem to result in a many-to-many mapping between psychological and neural categories.

1. A taxonomy of taxonomic positions

Before outlining what I take to be the main positions in this debate, I should point out that the taxonomic positions to be discussed are not mutually exclusive, in the sense that the same theorist could be a contextualist about psychological construct P_1 , and a revisionist about a different psychological construct P_2 (cf. McCaffrey 2015).

1.1. Localism

The first position regarding the neural–psychological structure-to-function mapping is the most familiar (and perhaps most popular). Localism has it that there will be a one-to-one mapping between brain regions and psychological or cognitive constructs. This is different from reductionism since it takes no stand on whether the correspondence will be type–type or token–token, though many localists think that correlations between cognitive functions and neural regions will hold for types not just tokens.¹ Moreover, like the other taxonomic positions to be discussed, it does not say whether the mapping or correlation is one of identity, causation, epiphenomenal emergence, or some other metaphysical relation. All it says is that the correlation between psychological and neural constructs will be such that the manifestation of a particular psychological construct will correspond to some localized neural state (process, event, and so on). This position need not require that the neural state be localized at a *single* brain region or location, for that would be unduly restrictive, and would seem to apply, in practice, to few if any neural and psychological states, given our current empirical knowledge. Instead, localism can allow that a set of regions (N_1, N_2, \dots, N_n), or indeed a network of regions, can be correlated with the manifestation of a psychological construct (P_1). In case this seems so weak as to be vacuous, it is worth emphasizing what localism would rule out. I take it that it would be incompatible with localism for some region (N_1) to be correlated, either singly or in conjunction with other brain regions, with two entirely distinct psychological constructs, P_1 and P_2 . The reason for this is quite straightforward. The very point of localism is that neural regions or networks are *maximally sensitive* or *selective*, dedicated to specific psychological tasks or functions, and that they are not multivalent, corresponding to distinct psychological constructs in different contexts or on different occasions. Having said that, a little care needs to be taken in saying what it means for two psychological constructs to be entirely distinct. If one psychological construct is a subordinate or superordinate category of another, then they are not entirely distinct constructs. For instance, if a neural region is correlated with the psychological construct *anger*, this clearly does not rule out the possibility that the same region is correlated with the superordinate category,

emotion. Indeed, that is just what we would expect. Superordinate and subordinate constructs are not entirely distinct. Similarly, overlapping psychological constructs are not entirely distinct, and localists would expect them to be correlated with overlapping brain regions or neural networks. What exactly it means for two psychological constructs to be overlapping is not entirely clear, but an example or two might help. If we think of *language* as a psychological construct, then *language comprehension* would be an overlapping construct, and *language comprehension* is likewise an overlapping construct with the psychological construct, *reading*. In each case, the latter function subsumes or presupposes the former and is hence not entirely distinct from it. The individuation of psychological constructs proceeds primarily according to their functional characterization, and even though it may not be possible to characterize the relevant notion of overlapping functions in purely general terms, I think the idea of one function subsuming another is familiar from exercises in “functional decomposition” (Bechtel and Mundale 1999, 190–201), as in constructing flowcharts or box-and-arrow diagrams. Hence, allowing for some vagueness in the notion of distinctness at play, I will take localism to say that the manifestation of a psychological construct (P_1) corresponds to the activity of a specific set of brain regions (N_1, N_2, \dots, N_n), provided that those same brain regions do not also correspond to the manifestation of some *distinct* psychological construct (P_2).

1.2. Globalism

A second position on the structure-to-function mapping is what I will refer to as globalism. This position simply says that the correlation between (at least some) psychological states and neural states cannot be localized in the vicinity of specific regions or networks, but implicate the whole brain, or at least large enough swathes of the cortex that it does not make sense to speak of regional or network correlations. The main idea here is that it is not primarily brain regions or neural networks that are identifiable with the different functional features of the brain, but rather such features as levels of activation and other global properties. Though there may be few if any researchers who are across-the-board globalists when it comes to psychological constructs, it is possible to find various examples of proposals for identifying specific psychological constructs with global rather than local neural states. For instance, there is a well-known claim due to Crick and Koch (1990) that consciousness is a “temporary global unity” imposed on neurons in many different parts of the brain by oscillations in 40–70 Hz range. Similarly, Suppes et al. (1999) associate perceptual concepts (e.g. *blue*, *yellow*, *line*, *triangle*) with patterns of brain wave activation (as indicated by EEG readings) over the entire cortex. Moreover, some versions of Dynamical Systems Theory as applied to neuroscience may posit a more thoroughgoing globalism for a broad range of psychological constructs.² As mentioned earlier, despite appearances, localism and globalism need not be mutually exclusive positions, in the sense that one could be a localist about some psychological constructs and a globalist about others. If one is not a reductionist, one might even be both localist and globalist about different tokens of the same *type* of psychological construct (say, in different individuals or different species), though one could not consistently be both about the same *token* psychological construct.

1.3. Revisionism

A third position, which has gained prominence recently, especially among neuroscientists, is what could be called “revisionism” (see, e.g. Price and Friston 2005; Poldrack, Halchenko, and Hanson 2009; Anderson 2015). As the label implies, this position calls for a

revision of our current psychological categories to align with neural constructs. It holds that the reason that there have been relatively few cases in which we have discovered a neat correspondence between neural and psychological constructs is that we have been incorrectly individuating psychological constructs. Accordingly, we will need to revise our psychological constructs if we are to discover neat structure-to-function mappings (whether of the local or global variety) between neuroscience and psychology.

Just as localism may seem, at first, to align naturally with a reductionist ontological account, revisionism may appear to be a natural ally of eliminativism.³ But here again, appearances are deceptive, since revisionists deny that our *current* psychological constructs correspond neatly to neural constructs (whether locally or globally), not that an amended system of psychological categories or constructs will do so. Of course, if the new taxonomic system is revised to the point that it bears little or no resemblance to our current system of categories, particularly those categories inherited from the folk, then the result might still be considered a version of eliminativism (at least with regard to folk psychological categories). But many revisionist proposals entail a reworking of our current constructs rather than a wholesale conceptual revolution. For example, Price and Friston (2005) suggest that activation in the left posterior lateral fusiform area performs the function of “sensorimotor integration.” In addition, Lindquist et al. (2012) speculate that the function of the amygdala is to process novel or motivationally salient stimuli rather than fear stimuli. Though these psychological constructs are perhaps not ones regularly invoked in psychological explanations (much less those of the folk) they seem like close relatives of some of our current scientific psychological constructs. If our current constructs were to be replaced by their close relatives, then we could be said to have revised them without having eliminated the psychological domain or undertaken a full-scale ontological displacement. Hence, taxonomic revisionists need not be ontological eliminativists. It should also be pointed out that one could combine revisionism with one of the other positions in this debate, in the sense that once one has revised the psychological categories, the resultant categories may then be mapped locally, globally, or contextually onto neural constructs.

1.4. Contextualism

Finally, the position that I will dwell on, since it is the one I will be arguing for, is what I will call, for lack of a better label, “contextualism.” In at least one of its variants, “neural contextualism,” this position is fairly well attested in the recent neuroscientific literature.⁴ But I will also be identifying another variant, “environmental-etiological contextualism,” which has not received its fair share of attention.

One way of introducing neural contextualism is by taking it to be a denial of one tenet of localism. In characterizing localism, I argued that it could not tolerate some neural region or network N_1 corresponding to two distinct psychological constructs, P_1 and P_2 . That is precisely what neural contextualism maintains can be the case (and sometimes is the case). To give a very simple schematic example of the kind of situation envisaged by this position, neural contextualism says it may be the case that activation in neural region N_1 , along with another region N_2 , jointly corresponds to psychological construct P_1 , but that activation in N_1 along with some other region N_3 corresponds to a distinct psychological state P_2 . To put it differently, a given neural region could be a part of two distinct neural networks, each of which performs distinct psychological functions. Moreover, N_1 , which is the common denominator among these networks, cannot by itself be said to perform an isolable psychological function that is a sub-function of the function performed by the entire network. There may seem to be something puzzling about this view, since it may

not be obvious how the very same region N_1 could play a distinct role depending on its set of neural alliances, or its overall neural *context*. Surely, we are tempted to say, there must be something distinctive that N_1 contributes, and this contribution should be capable of being extracted from or distilled out of the overall function of the network of which it forms a part.

There are two ways of responding to this worry, one of which proceeds by analogy and the other is more direct. First, consider the analogy of elemental atoms combining into molecules. When one atom of sodium bonds with one atom of chlorine to form a molecule of common salt, NaCl, the resulting molecule has microscopic and macroscopic properties that differ considerably from the properties of the molecule that combines a sodium atom with one hydrogen and one oxygen atom, sodium hydroxide or NaOH. Moreover, one cannot point to certain common properties of the two resultant molecules that correspond to the contribution of the sodium atom in particular. It is not the case that all sodium compounds have something in common with respect to their macro-properties or phenomenological features that one can trace back to the presence of sodium.⁵

This analogy may not be very convincing since it pertains to such a disparate scientific domain, but a direct reply may be constructed by attending more closely to one specific aspect of the analogy. To over-simplify, what is crucial in the case of the two molecules is the bonds between the atoms. Since sodium atoms bond differently with chlorine atoms than they do with oxygen and hydrogen atoms, we should not expect that the sodium atoms would make the very same contribution to the resulting molecules. In an analogous fashion, what is crucial in the case of brain regions is often not mere activation or co-activation but rather the precise nature of the anatomical and functional connectivity between the regions. Depending on the quantitative and qualitative properties of the connections between regions, we might expect that the contribution of some particular neural region will be different in different neural contexts. Some regions are connected via a dense network of neural connections while others are only sparsely connected. Some are related by feed-forward neural connections while others are connected via feedback or recurrent loops (or both). Moreover, interactions between brain regions occur over a period of time, and depending on the exact contribution of each region in a temporal sequence, the overall causal effects may be different. Hence, it certainly seems possible that activation in a given neural region could correspond to distinct psychological functions, depending on which other regions it is co-active with, based on the nature of the neural pathways that connect regions, their patterns of connectivity, the type of coupling that obtains between them, and the temporal dimensions of that coupling.

The theoretical possibility of neural contextualism is therefore not as mysterious as it might seem at first, and the empirical evidence for it is considerable (see, e.g. Anderson 2010, 245–246, and references therein). But there is another type of contextualism that pertains to the broader non-neural context, which does not seem to have been sufficiently explored. Neural networks are not activated in a vacuum and the broader context in which they are active might in some instances result in their subserving different functions. Even though the narrow functions they perform may be the same, the psychological and cognitive sciences are often interested in broader functions. As such, they are likely to individuate psychological processes and states with reference to their environmental context or causal history, at least on some occasions. In a certain respect, this point should be familiar from discussions of externalism in the philosophy of mind, but the case is sometimes made, in response, that externalistic individuation is not or ought not be the business of a science of psychology, and that as far as psychological explanation and prediction are concerned, all that matters is the narrowly individuated states of the organism. In other words, something like the attitude of “methodological solipsism” is sometimes recommended when it comes

to the scientific study of the mind and mental processes (Fodor 1980). But I will argue in what follows that for many strictly scientific purposes, there are clear cases in which individuation of psychological states and processes according to etiological or environmental considerations is indicated by scientific practice itself. Before putting forward some examples to make this case more convincing in Section 3, I will look more closely at the question of structure-to-function mapping, particularly as it arises in other scientific contexts.

2. Structure-to-function mapping

The idea of a structure-to-function mapping has become commonplace in neuroscience and cognitive science, but the notions of structure and function are typically left unanalyzed in the empirical literature. While the notion of function has been a subject of intensive analysis in philosophy, and there are several ways of understanding it, the notion of structure has not been as keenly discussed among philosophers, though it is not as transparent as it often appears. Moreover, the precise ways in which structure and function differ are often left unaddressed. Although it may not be possible to distinguish them without ambiguity and across all contexts, I will try to make the case that there are certain hallmarks of structural descriptions and explanations that tend to set them apart from functional ones. I will also try to delineate what I mean by structure and function when it comes to the mind–brain, and will argue that individuation according to structure and function do not always coincide, in the sense that the structure-to-function mapping can be many-to-many rather than one-to-one, or even many-to-one (as in multiple realization).⁶

2.1. Characterizing structure and function

One means of differentiating structural descriptions and explanations from functional ones is by way of characterizing the types of features or properties that are regularly invoked by each, though there may be no surefire way of distinguishing them or of specifying necessary and sufficient conditions that pertain to each. In attempting to characterize them, we should bear in mind that “structural” and “functional” may just be loose labels for explanations that track different types of causal patterns or trajectories.

Structural descriptions and explanations tend to emphasize intrinsic, synchronic, and spatial properties. That is to say, they often appeal to properties that pertain to the phenomenon that is being classified in isolation from other entities, or only in relation to contiguous or proximate other entities. They also typically cite properties that consider the phenomenon synchronically or at a given time, not taking into account its causal history or etiology (though this will be qualified in what follows). In addition, structural descriptions and explanations sometimes involve spatial features or refer to such properties such as length, distance, shape, size, dimensions, and so on. In this respect, structural descriptions and explanations are closely related to, though not identical with, mechanistic ones, which also often depend crucially on spatial relations.⁷ When it comes to the brain, neuroscientists sometimes distinguish structural descriptions at three different scales or levels, the micro-, macro-, and mesoscales. At the microscale, structural descriptions refer to individual neurons and their parts (e.g. axons, dendrites), at the macro-scale, they refer to brain regions and local networks (e.g. lobes, sulci, gyri), and at the mesoscale, they may pick out large-scale networks and “mini-columns” that span all cortical areas (Sporns, Tononi, and Kötter 2005). For the purposes of a structure-to-function mapping, it is often assumed that the microscale will be less relevant, so the macro- and mesoscales are the

focus of most of the research on mapping, and it is these levels of structure that I will assume are at play in this discussion. It should be acknowledged that at both the macro- and mesoscale, brain structures are often characterized in relational or extrinsic rather than intrinsic terms. But the relational properties are generally internal to the brain and do not involve extra-cranial structures. Though some ways of identifying neural locations at the macro- and mesoscales are clearly relational, the relations involve other regions of the brain or overall location within the brain as a whole. They do not typically involve relations to the environment or the broader context in which the organism is situated.

As for functional descriptions and explanations, they tend to abstract away from spatial properties and features, they often invoke extrinsic or relational properties, and they sometimes also advert to diachronic features of the phenomena to be explained, including etiology or causal history. There are two prominent philosophical accounts of function, or two “paradigms” of functional explanation in the biological and psychological domains (cf. Millikan 2002). The first, often associated with Millikan (1984), is particularly applicable to the biological realm (and by extension, the psychological) and it emphasizes the history of selection of a trait or phenotypic feature. This analysis characterizes the “proper function” of a trait as the causal effect that it was selected for (by the process of natural selection). Millikan emphasizes that there is no such thing as a *unique* proper function and that traits or “biological devices” often have both proximal and distal proper functions. But on her account, they are generally identified in terms of their etiology or causal history. The second family of accounts is usually traced to Cummins (1975), who characterizes the function of a capacity or device not in terms of a history of selection but rather in terms of its causal contribution to a broader system. As Millikan (2002, 118) summarizes it, Cummins’ notion of function is designed to clarify the kind of functional explanation that “explains how the organism’s system works, rather than why the traits contributing to these capacities or workings are there.” One of Cummins’ illustrative examples is a factory assembly line, in which the function of each worker or machine is to complete a certain task that contributes to the production of a commodity on that assembly line (1975, 760). While both accounts of function are relational, on the second account of function, the relations are to the system or broader synchronic context, whereas on the first, they are a matter of diachronic selection history. Sometimes the two accounts ascribe the same function to the very same device or trait, but at other times they diverge, as when a biological trait has been selected for one causal effect but comes to contribute to a different effect, perhaps in a different environmental context. (Of course, if the trait is later selected for the second causal effect, then it comes to fulfill that same function in the second sense as well. This is what is sometimes termed an “exaptation” or coopted adaptation, as when bird feathers were originally selected for insulation, later contributed to flight, and then were selected for this new role.) The first account of function is explicitly diachronic and etiological, adverting to a history of selection, while the second account is also extrinsic or relational, albeit with reference to a synchronic environmental or organismic context.⁸

When it comes to a system like the mind–brain, functions can be identified at multiple levels, but in discussing a “structure-to-function” mapping, many neuroscientists, particularly cognitive neuroscientists, have in mind a mapping to psychological or cognitive functions. In this context, they are usually not concerned with the neural functions of structural components at the microscale, such as the functions of ion channels or synapses in neurons – though there are many challenging and interesting questions about the precise functions of such neural structures. Rather, in this area of research, the focus is on a

mapping to broadly psychological functions. But even when it comes to psychological or cognitive functions, one can distinguish “lower-level” functions like automatic reflexes, from “higher-level” ones, such as language or emotion. Usually, the functions of interest are somewhat *more specific* higher-level functions, such as facial recognition, theory of mind, or reading comprehension. It is these, fairly circumscribed and specific psychological or cognitive functions, that cognitive neuroscientists and others are typically interested in mapping to neural structures. It may not be possible to spell out what constitutes a psychological function in any detail, but the general idea should be clear and these examples should help to delineate the types of phenomena that are the main objects of interest in the context of a structure-to-function mapping. Are these functions in the Millikan or Cummins sense? They are often both, since many human psychological capacities have been selected for, though they can also be said to make a causal contribution to a broader system, namely the organism situated in particular physical or social environments. However, it is also possible, particularly in the human case, that some psychological capacities have functions that were not originally selected for, since they have been adapted to perform those functions over the course of human history. This may be the case, for example, for psychological capacities like reading comprehension or chess-playing ability.

2.2. Mapping structure-to-function

Now that we have a rough-and-ready characterization of structure and function, we can look more closely at what is meant by a structure-to-function mapping. As we shall see, the differences between structure and function are not hard and fast, and there are some categories and taxonomies that combine structural and functional elements or features. Perhaps the most productive way to proceed is by looking at a hypothetical (and highly simplified) example. Imagine a biologist undertaking a structure-to-function mapping among limbs and locomotive activities. The task may seem trivial because the structures are so easily identifiable and their functions can (often) be read off of them, at least after observing the organism’s behavior for some time. But when faced with hitherto unobserved species, it can pose something of a challenge, and indeed, even when it comes to familiar species the “proper function” of a structure may not always be so obvious. (To be sure, ascertaining Cummins functions is usually more clear-cut than discovering Millikan functions, since the latter task involves knowing something about selection history, which is often difficult to reconstruct.) Imagine that our biologist is confronted for the first time with the forelimbs of various organisms from different mammalian species, say a dolphin, cat, and bat. The term “forelimb” seems like a straightforward structural one that pertains to the organ’s intrinsic properties, such as the number and arrangement of bones, as well as the organ’s position relative to the rest of the body, and so on. The biologist would soon notice that despite the structural similarities, the forelimb performs a different locomotive function in each of the organisms mentioned, namely swimming, walking, and flying, respectively. This is a case in which the *same* structure (in one sense to be elaborated), the mammalian forelimb, performs *multiple* functions across different species. Thus, it would appear as though there is a one-to-many mapping from structure to function. But it is also the case that each of these same functions can be multiply realized in different structures. For example, the function of flying is also performed by insects, whose wing structures are distinct from mammalian forelimbs, lacking many of their intrinsic properties.⁹ Therefore, it is not difficult to find biological cases in which there is a many-to-many relationship between structural and functional

taxonomies. Another way of putting this is to say that structural and functional taxonomies crosscut one another.

Before going any further, there are a few features of this apparently simple case that need to be brought out. First, in saying that the same structure (forelimb) performs multiple functions, we may be accused of ignoring the very significant differences among these structures. The forelimbs of cats, dolphins, and bats have different dimensions, detailed shapes, and so on, and these differences may be said to outweigh any structural similarities. Although it is true that the differences are crucial, this does not preclude classifying these structures under the overarching term or superordinate category “forelimb,” and this classification is not just a loose grouping of diverse organs but is grounded in important similarities. This leads to a second complicating feature of the example, which is that the category “forelimb” as it is being used here is not a purely structural designation, since it combines aspects of structure (number and arrangement of bones, and so on) with an etiological dimension that pertains to the common origin of this structure in these different organisms. The forelimbs of mammals are classified together partly on the basis of their descent from a common ancestral form; in other words, they are homologues. The commonalities among these organs do not have to do simply with their intrinsic properties and synchronic spatial properties but also with their common history of descent. “Forelimb” is therefore something of a hybrid category, combining as it does both structural and etiological features.¹⁰ The third distinctive feature of the example is that function in this case can be interpreted according to either Millikan’s or Cummins’ analyses, and they broadly coincide, since the functions mentioned (swimming, walking, flying) were both selected for and are also the functions that those forelimbs help to perform for these present-day organisms. Moreover, the history of selection that is brought out by (this particular way of) specifying the Millikan function differs from the history of selection that was just used to classify these organs as homologous forelimbs. Hence, even though there is an element of etiology in the superordinate category “forelimb” it refers to a more remote history of selection than the one that pertains to their different locomotive “proper” functions. A fourth complication is that there are narrow and wide ways of individuating functions, so it may be open to someone to say that the different functions identified above are not really different if one construes function broadly, say *locomotion*, which would cover swimming, walking, and flying. But the point is that one *can* individuate functions more narrowly, and provided this narrow mode of individuation identifies a function that is explanatorily significant, this mode of individuation would seem to be warranted.¹¹

Despite these complications, this case illustrates how different individuating practices can result in different taxonomies, ones which emphasize structure over function, diachronic over synchronic properties, and so on. Even though in this particular case both structural and functional designations have an etiological element to them, we still have a many-to-many mapping between structure and function, since they advert to different causal histories. If we grant that there is a single structural category, *forelimb*, that these phenotypic features belong to and that there are multiple functions performed by these organs, then we have a case of a many-to-many structure-to-function mapping when it comes to these morphological and behavioral phenotypes. The same structure (forelimb) can give rise to a multiplicity of functions (swimming, walking, flying), and the same function (flying) can be subserved by more than one structure (mammalian forelimb, insect wing). In other words, these structural and functional categories belong to crosscutting taxonomies (cf. Weiskopf 2011, 329; Stinson 2016, 1609).

3. Episodic memory as a case study

Episodic memory provides an instructive case study for a structure-to-function mapping in the mind–brain sciences. It is the subject of a lively area of research, for both cognitive psychologists and neuroscientists, and has been the focus of a flurry of recent empirical work. Often glossed as autobiographical memory, or memory for specific episodes in the subject’s own life, episodic memory has been researched over the past several decades using the methods of cognitive psychology, and more recently, using neuroscientific methods. While many research questions regarding this psychological construct are still unresolved, a number of recent findings have yielded some surprising philosophical lessons. In what follows, I will be arguing that there are some features of episodic memory that lend themselves to environmental-etiological contextualism, and hence suggest a many-to-many mapping between the psychological functions and neural structures, for some of the same reasons encountered in the previous section.

3.1. *The causal condition*

There has been an intriguing convergence of evidence from disparate sources indicating that episodic memories are more labile than has traditionally been assumed, that the act of recall often alters and transforms memories, and that there may be nothing akin to a definite memory trace or stored mental representation (sometimes called an “engram”) corresponding to each episodic memory. According to a body of work that is sometimes loosely labeled “Constructivism” about memory, memories are not discrete representations of particular past events. Moreover, remembering is more a matter of reconstruction than simple retrieval and incorporates background information as well as elements of the cue or context (cf. Nader and Hardt 2009; Robins 2016b).

Until fairly recently, much philosophical work on memory has been relatively impervious to empirical research findings and the field has been dominated by the “causal theory of memory” (Martin and Deutscher 1966). This theory says roughly that whether a psychological state is a memory and what it is a memory of are largely determined by a causal condition. More specifically, when it comes to a memory of a specific episode in an agent’s past, there needs to be a causal connection between an initial representation of the episode and a subsequent representation, and that causal connection must be mediated by a memory trace (cf. Michaelian 2011; Robins 2016a). In light of a preponderance of evidence indicating that memories are in fact much less stable than philosophers have tended to assume, that they are imperfectly encoded and stored, and that they are modified during recall, many proponents of the causal theory have altered their position. Rather than insist that the entirety of a representation be traceable continuously along a causal chain, via a memory trace, to the earlier representation of a specific episode, for it to qualify as a memory of that episode, many proponents of the causal theory allow for a looseness of fit between the earlier representation, the memory trace, and the subsequent representation. For instance, according to Michaelian (2011, 335), the memory trace must have a content that is “sufficiently close” to the content of the earlier representation. As Michaelian (2016, 92) puts it,

the updated version of the causal theory goes beyond available theories primarily in terms of its tolerance of divergences between the contents of experience and the contents of retrieved memories – that is, in terms of the extent to which it views memory as a generative source of information.

This accommodates some elements of constructivism and allows for the content of the memory trace to diverge from that of the initial representation, though how much divergence is left vague. This vagueness may not be a defect of the analysis, but rather a feature of the category of episodic memory.

Meanwhile, some philosophers have given up entirely on the causal condition on episodic memory, arguing that insistence on retaining this condition is not in keeping with empirical results. For example, Michaelian (2016, 60)¹² holds a simulational view of memory, according to which:

remembering is generative not preservative: it is not a matter of preserving a representation but rather of constructing, on the basis of stored information originating in a variety of different sources, as well as information available in the subject's current environment, a *new* representation of a past episode. In short, remembering is a matter of imagining or simulating the past. (emphasis in the original)

The idea that remembering is an act of imagining the past, not in principle different from imagining the future, is also one that has gained traction among some psychologists and neuroscientists, who think of both capacities as a matter of “mental time travel.”

If this simulational view of memory is correct, then there need not be a causal link between an earlier representation and a later representation, via a memory trace, for the later representation to be a memory. But this view effectively erases the distinction commonly made between memories and imaginings of the past (e.g. my memory of my own 10th birthday party as opposed to my imagining of my mother's 10th birthday party). To this, it might be said: so much for the common conception of memory. If our ordinary understanding of what a memory is, as elaborated by successive versions of the causal theory, is out of step with current empirical findings, then on a naturalist conception of philosophy, we should modify our philosophical theory to fit the empirical facts. While this line of argument is quite compelling, there are reasons to resist it from within a naturalist framework, and to insist on individuating memories based on causal history.

3.2. *Rehabilitating the causal condition*

Jettisoning the causal condition on episodic memory deprives us of a natural way of distinguishing remembering and misremembering. However, Robins (2016a) argues persuasively that the abandonment of memory traces (whether manifest or dispositional) is an over-reaction. Instead, she makes the case that misremembering is a distinctive type of memory error, one that is explicable only on the assumption that the subject has retained information from the event that the representation characterizes. On this view, (episodic) misremembering is different from confabulation, since confabulations reflect no influence of information retained from a particular past event, whereas misrememberings do. It is important to tread carefully here, since there are at least two dimensions along which one might distinguish these three types of representation: remembering, misremembering, and confabulation. The first is causal origin and the second veridicality or accuracy. When it comes to causal origin, I take it that there is a clear distinction between the first two types of representation and the third. While remembering and misremembering both have their origin in a representation of a past episode, confabulation does not.

On this way of doing things, the difference between remembering and misremembering does not reside in their respective degree of accuracy, but has to do with the extent to which the representation has been adulterated or altered. Misrememberings are representations

that are causally linked to representations of the past but have subsequently been distorted with additions and deletions. As for accuracy, it does not correlate neatly with causal origin. It may well be that a representation is accurate but does not originate either wholly or partly in a past episode; conversely, a representation may be thoroughly inaccurate yet be linked to a past episode. If it departs sufficiently from the original representation, a representation can be considered a misremembering even though the distortions make it more accurate as a result (say, in case the original representation was non-veridical or incomplete¹³). Of course, this means that the difference between remembering and misremembering is a matter of degree, with remembering being more tightly linked to the representation of a past episode and misremembering being linked only in an attenuated and diluted fashion. The difference between them depends on the extent to which their content is causally dependent on the content of a representation that originates in a past episode. Confabulation, on the other hand, is not on this continuum at all, since it is distinguished from both remembering and misremembering in not being causally linked to a representation of a past episode at all. Here, I depart somewhat from Robins (2016a, 434), who defines misremembering as follows:

Misremembering is a memory error that relies on successful retention of the targeted event. When a person misremembers, her report is inaccurate, yet this inaccuracy is explicable only on the assumption that she has retained information from the event her representation mischaracterizes.

In this definition, the question of accuracy seems to be conjoined with the question of origin (successful retention), whereas I am advocating disentangling the two dimensions. Robins (2016a, 434) also writes: “The difference between misremembering and confabulation may, ultimately, be one of degree.” Rather, I will posit that the difference between misremembering and confabulation has to do with whether or not the representation originates in a past episode and retains information from a prior representation. This is an on–off distinction, by contrast with the distinction between remembering and misremembering, which is indeed a matter of degree, namely the degree to which the representation retains information from the original representation of the past episode (which may itself be inaccurate or seriously incomplete).

Why should we want to insist on a threefold distinction between remembering, misremembering, and confabulation, made on the basis of causal origin (rather than accuracy)? There are a couple of reasons why a cognitive psychologist would be interested in asserting the distinction between remembering, misremembering, and confabulating on an etiological basis. For one thing, this squares with a common explanation of one paradigmatic type of confabulation. In Korsakoff’s syndrome, patients often engage in confabulation due to a failure of episodic memory. On one explanation, in Korsakoff’s the episodic memory system is isolated from the conceptual and linguistic systems, which are responsible for issuing verbal reports. As Hirstein (2000, S423) puts it: “confabulation seems to occur when analog representation systems [including episodic memory] are disturbed or isolated from the conceptual representation system in the left hemisphere.” Since the memory trace is not causally connected to the formation of a conceptual and linguistic representation, the causal chain linking the initial representation to the subject’s current representation is disrupted. On this theory, it is this lack of a link to a past episode via an initial representation, rather than lack of accuracy, which characterizes confabulations and sets them apart from memories. Of course, that is not the only way of understanding confabulations, but a number of other theories of confabulation conceive of them as involving “deficits in

strategic memory retrieval or disrupted storage of (or access to) autobiographical memories” (Bortolotti and Cox 2009, 955). Meanwhile, very few theories define confabulation in terms of lack of accuracy, and those that do are open to the objection that some confabulations happen to be accurate or true (cf. Bortolotti and Cox 2009, 956). Thus, the proposed distinction between remembering and misremembering (on the one hand) and confabulating (on the other) aligns very well with explanations of confabulation, which account for them by outlining the way in which the causal connection between initial representation and current representation is disrupted.

A second reason for distinguishing the three types of representation (remembering, misremembering, and confabulating) on an etiological basis has to do with ascertaining the function of memory. The distinction between remembering and misremembering is a presupposition of many of the paradigmatic experimental studies on memory. In the Deese–Roediger–McDermott (DRM) paradigm, subjects are presented with a list of items and asked to recall as many as possible after a time interval, or asked to recognize them from another list. The main finding is that subjects often spontaneously come up with or claim to recognize items not on the list, especially if those items are thematically related. For example, after being presented with a list containing words like “bed,” “rest,” and “awake,” among others, participants will later mistakenly say that the list also contained the word “sleep” (Roediger and McDermott 1995). The point of these experiments is to compare the items that were on the list to the items that were “recalled” or “recognized” but not on the list, in order to determine how episodic memory modifies the initial representation of the presented list. The whole exercise is undertaken against a backdrop of a presumed distinction between remembering and misremembering. The reason is not hard to find. If cognitive psychologists are interested in ecological validity, then they will be interested in understanding the circumstances and mechanisms that differentiate the formation of these two kinds of representation, and the ways in which the processes of storage and recall can be disrupted.¹⁴ Yet, in terms of the properties of the mental representations involved, there may be no difference between the intrinsic properties of remembering and misremembering. The only difference between them is the causal chain that led to their manifestation. In other words, diachronic or etiological factors are what distinguish them, and these factors are not inherent in the representation itself, much less in its neural counterpart.

Distinguishing remembering from misremembering is crucial to discerning the function of memory. Some constructivists have argued that the function of memory is not to faithfully record past experiences but rather to reconstruct these experiences. Michaelian (2016, 99) sums up the emerging consensus on the function of episodic memory as follows: “The standard view, in short, is increasingly that episodic memory is one function of a more general episodic construction system, a process not different in kind from imagining a range of nonactual episodes.” To be sure, such a system may be more accurate and reliable than one that is designed merely to preserve past representations without modification, drawing as it does on reliable background information, since past representations may not have been accurate in the first place and storing them faithfully may not be cost-effective. But even if these theorists are right about this, the discussion rests on the ability to make a distinction between a system whose function is to faithfully encode, store, and retrieve representations of past events and one with an altogether different function. Any investigation of this question needs to make a distinction between remembering and misremembering. Having said that, it should be admitted that cognitive psychologists investigating episodic memory are not in a good position to differentiate remembering from misremembering directly on the basis of causal origin. For example, in the DRM paradigm, it is not clear how investigators can distinguish between instances in which participants

correctly recall “awake” from a list and those in which they just happen to come up with it (much as they come up with “sleep”). Even though participants are often instructed not to guess, it is practically impossible to determine whether a participant who recalls the word “awake” is actually relying on a memory trace of that word or is inferring its presence on the list due to memory traces of other words on the list. Rather than try to determine the causal history of these representations, investigators rely on the surrogate of veridicality. Even though correct answers are not necessarily ones that faithfully reproduce a past representation, the assumption in this paradigm is that if there are enough trials, investigators can distinguish what we might consider “lucky guesses” from actual recall or recognition. Thus, cognitive psychologists do not have the tools to track causal origin directly, so they distinguish representations based on their veridicality.

I have been arguing that the causal condition is necessary to make a scientifically valuable distinction between remembering and misremembering in the psychological domain. I have also suggested that such a distinction is not observed in the neural domain, where individuation of neural states does not pay heed to distal causation and etiology, and this is why there is likely to be a mismatch between psychological and neural taxonomy. Just as in the case of anatomical or morphological features, structural and functional taxonomies can lead to a many-to-many mapping, we might expect that in the neural–psychological case we will also observe crosscutting taxonomies due to differing structural and functional individuating practices. Against this, it may be said that neuroscientists may indeed be capable of distinguishing among neural states and mechanisms based on their origin or causal history. To be sure, there is nothing to prevent them from distinguishing neural representations based on their causal history and individuating them on that basis. Neuroscientists could choose to classify neural states and mechanisms with reference to their distal causes, but at least given the methods and concerns of contemporary neuroscience, and the emphasis on structural and mechanistic explanations, there is reason to think that a structurally based taxonomy will predominate. The use of neural imaging and other techniques of investigating the neural mechanisms underlying episodic memory do not lend themselves to the isolation of the causal histories or origins of neural representations. Contemporary neuroscience, even at the macro- and mesoscales, does not appear to incorporate etiology in its taxonomic schemes.

I have argued that cognitive psychology taxonomizes, at least in part and on some occasions, on the basis of etiology or causal history. Meanwhile, neuroscience, with its current tools and methods, is not suited to this mode of individuation and is not oriented towards distinguishing states and mechanisms on the basis of their causal history. This is not to say that it might not develop such tools and reorient its methods of investigation and taxonomic practices so as to track etiology. Indeed, recent developments in optogenetics may lead to such a reorientation, since they afford neuroscientists the tools to control brain circuits with light and hence to plant “false memories” in animal models. In this experimental paradigm, genes are inserted in neurons that lead to the synthesis of light-responsive protein receptors (and these genes are typically coupled with genes that synthesize an imaging protein, for better detection). When stimulated by light, the protein receptors in the affected neurons can then lead to the activation of those neurons. This allows researchers to identify a group of neurons that are active during learning in a given context, and then to reactivate those neurons in a different context (Liu et al. 2012). Thus, neurons associated with a fearful stimulus (electric shock) are activated in the animals and associated with a context in which no fearful stimulus is received. The animals in this second context are found to exhibit fearful behavior, even though they had not previously experienced the fearful stimulus in this second context.¹⁵ This technique allows neuroscientists to more effectively track the distal causes of memory representations and to manipulate those causes, in such a way as to replace an original

representation with an alternative representation, effectively implanting a “false memory” (cf. Liu, Ramirez, and Tonegawa 2014). Here again, what is crucial is not so much lack of conformity to the original stimulus (objective truth or falsity) but the extent to which the original representation has been distorted by associating it with the representation of a fearful stimulus. If this and similar techniques become more widespread, they may enable neuroscientists to systematically distinguish neural traces on the basis of their provenance or causal origin. If they can do so reliably, they may come to taxonomize on that basis, thus aligning their categories more closely with those of cognitive psychologists, distinguishing remembering from misremembering roughly in the way that psychologists do. But even this does not guarantee an alignment in their respective taxonomic schemes, since as we saw in the case of morphology, etiology may enter into taxonomy in different ways, depending on whether one is focusing on remote common ancestry (homology) or a more recent selection history (proper function).¹⁶

4. Conclusion

Structure-to-function mappings can yield a many-to-many or crosscutting relationship between structural and functional taxonomic categories. It is not that structural and functional taxonomies cannot coincide, or that there can be no hybrid categories, it is just that whenever we attempt to map structure onto function, there is a possibility of a many-to-many mapping, given the different classificatory interests of the associated taxonomic schemes. In particular, where structures are individuated synchronically and functions diachronically, or when they are individuated with respect to different diachronic trajectories, as they often are, there should not be an expectation of a one-to-one, or even a one-to-many mapping between the different taxonomies. Instead, the categories will crosscut one another, thus preventing a type reduction of the kind favored by identity theorists. I have put forward some considerations to suggest that taxonomic categories associated with episodic memory (*remembering, misremembering*) are sensitive to distal environmental causes and to etiological considerations (environmental-etiological contextualism), whereas neuroscientific classification tends not to track contextual and etiological features. However, I have also allowed that the advent of new methods and experimental techniques may change the classificatory landscape. All claims of reducibility or lack thereof are contingent on the contemporaneous theories and methods of the domains involved. As Ernest Nagel (1961, 363) recognized, in his classic work on reductionism:

The irreducibility of one science to another (for example, of biology to physics) is sometimes asserted absolutely and without temporal qualifications. In any event, arguments for such claims often appear to forget that the sciences have a history, and that the reducibility (or irreducibility) of one science to another is contingent upon the specific theory employed by the latter discipline at some stated time.

It remains to be seen whether recent developments in neuroscience will issue in taxonomies that are more aligned with those of cognitive psychology, leaving open the possibility of reducibility at some future juncture.

Acknowledgements

I am very grateful to three anonymous referees whose constructive criticisms led to several improvements to this paper. I am also indebted to audiences at the conference on Rethinking the Taxonomy of Psychology, Western University, and the Department of Philosophy, University of Illinois, Chicago, whose feedback on earlier versions of this paper was very helpful in revising it.

Disclosure statement

No potential conflict of interest was reported by the author.

Notes

1. There are a number of distinct conceptions of reductionism, but I think that what I say here applies to any conception that understands reduction as a type–type relation.
2. Another version of globalism might invoke the theory of “neuronal packets” (Luczak, McNaughton, and Harris 2015), according to which the temporally organized coordinated activity of neuronal populations constitute the basic building blocks of cortical coding.
3. I understand eliminativism in a minimal sense to comprise any position that hypothesizes the elimination of all mental or psychological categories in favor of neural, physiological, or biological ones.
4. Lindquist et al. (2012) call a closely related position “constructionism,” and this position also seems close to what McIntosh (2004) terms the “neural context hypothesis.”
5. For other analogies in this vein, see Klein (2012, 955) on the changing functions of pistons in diesel truck engines, and Lindquist et al. (2012, 126) on the variable contributions of the same ingredient in different recipes.
6. An early and illuminating account of the difference between structure and function in this context is given by Fodor (1968, 113):

If I speak of a device as a “camshaft,” I am implicitly identifying it by reference to its physical structure, and so I am committed to the view that it exhibits a characteristic and specifiable decomposition into physical parts. But if I speak of the device as a “valve lifter,” I am identifying it by reference to its function and I therefore undertake no such commitment.

7. I think this close relation obtains given the two most widely cited definitions of mechanisms, which regard them either as “entities and activities organized such that they are productive of regular changes from start or set-up to finish or termination conditions” (Machamer, Darden, and Craver 2000, 3), or as “an organized set of parts that perform different operations which are orchestrated so as to realize in the appropriate context the phenomenon in question” (Bechtel 2009, 544).
8. One could also come up with a kind of hybrid account; for example, Griffiths (1993, 409) argues that the proper functions of a biological trait are “the functions it is assigned in a Cummins-style functional explanation of the fitness of ancestors.”
9. Wings are a good example of convergent evolution, since they arose independently in four different lineages in the history of life (mammals, birds, pterosaurs, and insects). But despite the fact that these analog structures evolved independently, they exhibit many similarities.

As far as we can tell, and despite their radically different starting points, the discovery of flapping flight by the ancestors of bats, pterosaurs, and insects probably played out in much the same way each time, with a controlled descent phase and a progressively refined gliding stage preceding the thrust-generating breakthrough, all ultimately fuelled by gravitational potential energy. (Wilkinson 2016, 70)

10. These homologous organs can be considered to correspond to the same character but to different character states (Brigandt 2007; Ereshefsky 2012).
11. See also Price and Friston (2005, 268) on narrow and wide construals of function. This discussion of function individuation brings up questions of what justifies the adoption of a particular system of taxonomic categories, and when such systems correspond to genuine natural kinds. Here, I am basing myself on an account I have given elsewhere (Khalidi 2013).
12. Michaelian altered his position from his (2011) to his (2016).
13. One systematic way in which some episodic memories are more complete than the corresponding original representations occurs in “boundary extension,” wherein subjects report seeing more of a visual scene than they actually saw (Michaelian 2011).

14. Again, truth and accuracy are not the main considerations, since misremembering can produce more accurate results in some circumstances (e.g. as in some instances of “boundary extension,” mentioned previously).
15. See Robins (2016b) for a more perspicuous and comprehensive description of this technique, including many further details. She argues that this experimental paradigm, which relies on the identification of a memory trace or engram, is incompatible with radical versions of constructivism.
16. Interestingly, Ereshefsky (2007) argues that treating psychological traits as homologues is a more promising theoretical approach than functionalist accounts or adaptationist accounts (which may track different etiologies).

Notes on contributor

Muhammad Ali Khalidi is Professor of Philosophy and Chair of the Department of Philosophy at York University in Toronto. He specializes in general issues in the philosophy of science (especially, natural kinds and reductionism) and philosophy of cognitive science (especially, innateness, concepts, and domain specificity). His book, *Natural Categories and Human Kinds*, was published by Cambridge University Press in 2013, and he has recently been working on topics in the philosophy of social science (especially, social ontology).

References

- Anderson, M. L. 2010. “Neural Reuse: A Fundamental Organizational Principle of the Brain.” *Behavioral and Brain Sciences* 33 (4): 245–266.
- Anderson, M. L. 2015. “Mining the Brain for a New Taxonomy of the Mind.” *Philosophy Compass* 10 (1): 68–77.
- Bechtel, W. 2009. “Looking Down, Around, and Up: Mechanistic Explanation in Psychology.” *Philosophical Psychology* 22: 543–564.
- Bechtel, W., and J. Mundale. 1999. “Multiple Realizability Revisited: Linking Cognitive and Neural States.” *Philosophy of Science* 66 (2): 175–207.
- Bortolotti, L., and R. E. Cox. 2009. “‘Faultless’ Ignorance: Strengths and Limitations of Epistemic Definitions of Confabulation.” *Consciousness and Cognition* 18 (4): 952–965.
- Brigandt, I. 2007. “Typology Now: Homology and Developmental Constraints Explain Evolvability.” *Biology & Philosophy* 22 (5): 709–725.
- Crick, F., and C. Koch. 1990. “Towards a Neurobiological Theory of Consciousness.” In *Seminars in the Neurosciences*, Vol. 2, 263–275. Philadelphia, PA: Saunders Scientific.
- Cummins, R. 1975. “Functional Analysis.” *Journal of Philosophy* 72 (20): 741–765.
- Ereshefsky, M. 2007. “Psychological Categories as Homologies: Lessons from Ethology.” *Biology & Philosophy* 22 (5): 659–674.
- Ereshefsky, M. 2012. “Homology Thinking.” *Biology & Philosophy* 27 (3): 381–400.
- Fodor, J. A. 1968. *Psychological Explanation*. New York: Random House.
- Fodor, J. A. 1980. “Methodological Solipsism Considered as a Research Strategy in Cognitive Psychology.” *Behavioral and Brain Sciences* 3 (1): 63–73.
- Griffiths, P. E. 1993. “Functional Analysis and Proper Functions.” *The British Journal for the Philosophy of Science* 44 (3): 409–422.
- Hirstein, W. 2000. “Self-deception and Confabulation.” *Philosophy of Science* 67: S418–S429.
- Khalidi, M. A. 2013. *Natural Categories and Human Kinds*. Cambridge: Cambridge University Press.
- Klein, C. 2012. “Cognitive Ontology and Region-Versus Network-oriented Analyses.” *Philosophy of Science* 79 (5): 952–960.
- Lindquist, K. A., T. D. Wager, H. Kober, E. Bliss-Moreau, and L. F. Barrett. 2012. “The Brain Basis of Emotion: A Meta-analytic Review.” *Behavioral and Brain Sciences* 35 (03): 121–143.
- Liu, X., S. Ramirez, P. T. Pang, C. B. Puryear, A. Govindarajan, K. Deisseroth, and S. Tonegawa. 2012. “Optogenetic Stimulation of a Hippocampal Engram Activates Fear Memory Recall.” *Nature* 484 (7394): 381–385.
- Liu, X., S. Ramirez, and S. Tonegawa. 2014. “Inception of a False Memory by Optogenetic Manipulation of a Hippocampal Memory Engram.” *Philosophical Transactions of the Royal Society B* 369: 2013–2042.

- Luczak, A., B. L. McNaughton, and K. D. Harris. 2015. "Packet-based Communication in the Cortex." *Nature Reviews Neuroscience* 16: 745–755.
- Machamer, P., L. Darden, and C. F. Craver. 2000. "Thinking About Mechanisms." *Philosophy of Science* 67: 1–25.
- Martin, C. B., and M. Deutscher. 1966. "Remembering." *The Philosophical Review* 75 (2): 161–196.
- McCaffrey, J. B. 2015. "The Brain's Heterogeneous Functional Landscape." *Philosophy of Science* 82: 1010–1022.
- McIntosh, A. R. 2004. "Contexts and Catalysts: A Resolution of the Localization and Integration of Function in the Brain." *Neuroinformatics* 2 (2): 175–182.
- Michaelian, K. 2011. "Generative Memory." *Philosophical Psychology* 24 (3): 323–342.
- Michaelian, K. 2016. *Mental Time Travel: Episodic Memory and Our Knowledge of the Personal Past*. Cambridge, MA: MIT Press.
- Millikan, R. 1984. *Language, Thought, and Other Biological Categories*. Cambridge, MA: MIT Press.
- Millikan, R. 2002. "Biofunctions: Two Paradigms." In *Functions: New Essays in the Philosophy of Psychology and Biology*, edited by A. Ariew, R. Cummins, and M. Perlman, 113–143. New York: Oxford University Press.
- Nader, K., and O. Hardt. 2009. "A Single Standard for Memory: The Case for Reconsolidation." *Nature Reviews Neuroscience* 10 (3): 224–234.
- Nagel, E. 1961. *The Structure of Science*. New York: Harcourt, Brace, and World.
- Poldrack, R. A., Y. O. Halchenko, and S. J. Hanson. 2009. "Decoding the Large-scale Structure of Brain Function by Classifying Mental States Across Individuals." *Psychological Science* 20 (11): 1364–1372.
- Price, C. J., and K. J. Friston. 2005. "Functional Ontologies for Cognition: The Systematic Definition of Structure and Function." *Cognitive Neuropsychology* 22 (3–4): 262–275.
- Robins, S. K. 2016a. "Misremembering." *Philosophical Psychology* 29 (3): 432–447.
- Robins, S. K. 2016b. "Optogenetics and the Mechanism of False Memory." *Synthese* 193 (5): 1561–1583.
- Roediger, H. L., and K. B. McDermott. 1995. "Creating False Memories: Remembering Words Not Presented in Lists." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 21 (4): 803–814.
- Sporns, O., G. Tononi, and R. Kötter. 2005. "The Human Connectome: A Structural Description of the Human Brain." *PLoS Computational Biology* 1 (4): e42.
- Stinson, C. 2016. "Mechanisms in Psychology: Ripping Nature at its Seams." *Synthese* 193: 1585–1614.
- Suppes, P., B. Han, J. Epelboim, and Z. L. Lu. 1999. "Invariance of Brain-wave Representations of Simple Visual Images and Their Names." *Proceedings of the National Academy of Sciences* 96 (25): 14658–14663.
- Weiskopf, D. A. 2011. "The Functional Unity of Special Science Kinds." *The British Journal for the Philosophy of Science* 62: 233–258.
- Wilkinson, M. 2016. *Restless Creatures: The Story of Life in Ten Movements*. London: Icon Books.