What Counts as a Memory? Definitions, hypotheses, and "kinding in progress"

David Colaço

Preprint

(Accepted for publication by *Philosophy of Science* on 11/17/2020)

Please cite Final Version

Abstract: This paper accounts for *broad definitions of memory*, which extend to paradigmatic memory phenomena, like episodic memory in humans, and phenomena in worms and sea snails. These definitions may seem too broad, suggesting that they extend to phenomena that don't count as memory or illustrate that memory is not a natural kind. However, these responses fail to consider a definition *as a hypothesis*. As opposed to construing definitions as expressing memory's properties, a definition as a hypothesis is the basis to test inferences about phenomena. A definition as a hypothesis is valuable when the "kinding" of phenomena is on-going.

Acknowledgements: Thanks to John Bickle, Eric Hochstein, Edouard Machery, and Sarah Robins for their comments on an earlier draft of this paper. This paper was presented at York University's "Natural Kinds in Cognitive Science" workshop as well as "Issues in Philosophy of Memory 2" at Université Grenoble Alps. Thank you to those who attended these presentations.

1. Introduction

The study of memory has recently seen new proposals for defining its target. Several researchers propose what I call *broad definitions* of memory, such as the idea that "memory can be defined as experience-dependent modification of internal structure, in a stimulus-specific manner that alters the way the system will respond to stimuli in the future as a function of its past" (Baluška and Levin 2016, 2). Others suggest defining memory by dropping the idea that the response is a specific function of the stimulus (Bédécarrats et al. 2018). These definitions extend to many phenomena, from episodic memory to reflex sensitization, some which occur in "unconventional cognitive systems" (Baluška and Levin 2016, 12). They extend beyond exemplars like me remembering what I ate for breakfast and do not express notions like representation or phenomenology that we often find in rich definitions that narrowly apply to kinds like declarative memory (Dudai 2007).

What might we expect out of responses to broad definitions? While there is no agreement about what counts as a memory, two potential responses are found in the existing literature. First, one might argue that these definitions are deficient. They extend to phenomena that "most of us would consider to fall outside the rubric of memory" (Schacter 2007, 24). Second, one might argue that these definitions illustrate that memory is not a natural kind, given the differences between the phenomena that comprise its members. If this is the case, then the "investigation of one kind of memory will not typically tell us anything about the other kinds of memory" (Michaelian 2010, 171). This response suggests that memory phenomena (broadly defined) do not share the relevant properties to support inferences about them.

These responses show that what counts as a memory is a controversial question for both philosophers and scientists. That being said, both responses are negative. They both suggest that

there is little if any scientific value in adopting a broad definition of memory, and they hint at the idea that the proposals of broad definitions of memory are best explained sociologically: they are metaphors or exaggerations that increase the research's perceived relevance and funding opportunities. This leads me to ask: why might researchers propose a broad definition of memory?

In this paper, I explain why researchers might adopt a broad definition of memory, which reflects a role for memory definitions that parallels debates about defining 'cognition' (Allen 2017) and 'life' (Bich and Green 2018). Broad definitions can promote the investigation of phenomena counted as memory on these definitions. This facilitates testing what claims about phenomena *project*. Thus, rather than expressing what is known about memory, researchers propose these definitions to test inferences about phenomena. In this sense, broad definitions of memory serve as *hypotheses*.

I proceed as follows. In *Section 2*, I discuss the scientific context of two broad definitions of memory. In *Section 3*, I introduce definitions as hypotheses. In *Section 4*, I address the *dissimilarity objection*: there are too many differences amongst phenomena for broad definitions of memory to be scientifically valuable. In *Section 5*, I address the *natural kinds objection*: generic memory is not a natural kind, so no definition of it is valuable. In *Section 6*, I address the *knowledge objection*: enough about memory is known, so there is no need for definitions of memory as hypotheses. Throughout, I argue that adopting a definition as a hypothesis exemplifies what I call *kinding in progress*: categorizing targets when their discovery is on-going, and their shared properties are largely unknown.

2. Historical and Scientific Context for Two Broad Definitions of Memory

The broad definitions of memory that I introduced – namely, experience-dependent modification of internal structure that affects stimulus response and may (or may not) be stimulus specific – stem in part from the study of the "regeneration" and the "transfer" of memory. This research has a complex history (see Colaço 2018). For this paper, I review two sets of experiments. Starting in the 1950s, McConnell studied whether planarian worms are candidate model organisms for associative learning. Upon purportedly demonstrating that planarians can associate stimuli (Thompson and McConnell 1955), he hypothesized that these associations would persist if the worms were cut into pieces and then allowed to regenerate (McConnell, Jacobson, and Kimble 1959). While McConnell seemingly confirmed that the regenerated worms can associate the stimuli, the research was controversial for two reasons. There were concerns with the design of his experiment, and there was debate over whether it counted as the regeneration of *memory*, as McConnell called it.

McConnell hypothesized that associations could be transferred from one organism to another. To investigate "memory transfer," McConnell designed an infamous study where planarians were trained to associate stimuli and then consumed by untrained cannibals, following which these cannibals' performance of behaviors consistent with the association of stimuli was measured (McConnell 1962). While a cottage industry developed from this controversial work, this research project collapsed by the 1970s. It has been argued that this collapse was due to replication failures and confounds in these studies, which undercut inferences that "positive" results were best explained by the occurrence of the transfer or regeneration of memory (Colaço 2018, 39).

There likely were experimental issues with McConnell's research, but the collapse of this research project also stemmed from the fact that his experiments were designed to detect what the researchers at the time thought was essential to studying memory. A common commitment of memory researchers of this era was that memory is associative (Colaço 2018, 34), and, because of this, McConnell characterized memory regeneration and memory transfer as associative phenomena. These associations were never reproduced by researchers, which supported the suspicion that, even if McConnell had identified novel phenomena, they do not count as memory phenomena.

Despite the research project collapsing in the 1970s, we have seen a sort of revival of it in the past decade. Shomrat and Levin substantially revised McConnell's 1959 design, and, in doing so, they putatively demonstrated the association of stimuli in regenerated planarians (Shomrat and Levin 2013). This finding has led behavioral researchers like Deochand and colleagues to suggest that "planarian regeneration is also of interest as to what facilitates memory or learning" (2018, 457). More recently, researchers from the laboratory of Glanzman report the "transplant" of sensitization via the transfer of RNA in aplysia sea snails, a model organism (Bédécarrats et al. 2018). This finding suggests that "RNA is sufficient to generate a priming component of the engram for long-term sensitization," paving the way for explaining this phenomenon (Carney 2018, 1). This research has run parallel with studies putatively demonstrating the intergenerational transfer of learned behaviors via RNA in C. elegans, another model organism (Moore et al. 2018). These findings suggest that these transfers may occur without human intervention (Posner et al. 2019).

Two features differentiate the contemporary from the historical research projects. First, the studies differ, which may resolve the experimental issues reported by Colaço (2018). This is

not a guarantee – the research may collapse again for experimental reasons – but this concern is not my focus. Second, and central to this paper, contemporary researchers propose broad definitions of memory. This is exemplified by the work from the Glanzman laboratory, in which non-associative sensitization is counted as memory. If one adopts one of these definitions, the phenomena that are discovered count as memory phenomena even if McConnell would not have counted them. This fact reveals the rationale for these definitions: by broadening what counts as a memory, researchers can count their discoveries as memory phenomena.

3. Definitions of Memory as Hypotheses

With the context of the proposal of broad definitions of memory introduced, we can investigate the function of these definitions. But, before I address these functions, there are three disclaimers that I must make. First, the broad definitions I introduced are not equivalent. One specifies stimulus specificity. The other does not. This difference affects what counts as a memory on the respective definitions, and it reflects the controversial research projects in which each originated. Baluška and Levin's definition links to research on attributing cognition to a wider set of biological and non-biological systems than those traditionally considered to be cognitive (2016). The definition from the Glanzman laboratory links to research on cross-organismic transfer phenomena and the function of RNA (Bédécarrats et al. 2018).

Second, accepting either definition is not equivalent to rejecting that there are distinct subtypes of memory. The broad definitions are of what might be called "generic memory" (Rupert 2004, 407) or memory as a superordinate category. Thus, they do not supplant characterizations of subtypes like declarative memory, nor do they entail that any phenomenon counts as (say) declarative memory if it counts as a memory. Even if one adopts a broad

definition, one can nonetheless accept that there are memory phenomena that are characterized differently. However, one commits to the appropriateness of this definition for each memory subtype.[1]

Third, both definitions I discuss have content: not everything counts as a memory according to either definition. Looking at the Baluška and Levin version, as it is explicitly stated, this definition includes stimulus-specific and experience-dependent change to the system's internal structure. Stimulus specificity and internal structure rule out Schacter's counterexample of breaking an arm as a memory, which Schacter presented to critique a broad definition of memory in terms of "experience that persists over time" (2007, 24).[2] Thus, Baluška and Levin's definition does not "fail to rule out much of anything," which is a complaint levied at past broad definitions (Schacter 2007, 24). However, this definition fails to rule out phenomena not traditionally considered to be memory, such as some perceptual and morphological phenomena. Lacking stimulus specificity, the definition implied by Glanzman's laboratory fails to rule out even more phenomena. This "failure" is deliberate: both research projects extend what counts as a memory.

Consequently, these definitions depart from much of mainstream memory research. They do not express the storage or retrieval of representations (Anderson 1990; Dudai 2007). Unlike philosophers (Michaelian 2010) and scientists (Tulving 2007) who distinguish declarative and non-declarative memory, these definitions apply to both subtypes and express nothing about distinct memory systems. They extend to phenomena in and beyond Squire's dominant

[1] Broad definitions of memory can be disentangled from literalism regarding statements about remembering. Because a broad definition of memory does not preclude there are distinct subtypes of memory, one need not endorse the idea that "the semantic contribution of [memory] predicates to truth conditions is the same in statements about the relevant nonhumans as it is in statements about humans" (Figdor 2018, 61).
[2] Even if arm breaking counted as a memory on Baluška and Levin's definition, the peripheral, external effects of the accident that resulted in the break would not count as part of the memory.

taxonomy of memory subtypes (2004). Additionally, these are just two examples of broad definitions of memory. Others are broader in some respects, such as those that drop commitments to internal structure.[3] These definitions undoubtedly play a rhetorical role: branding one's research as memory increases its prestige. This is a feature of broad definitions that should not be dismissed. However, this feature does not exhaust the importance of these definitions.

3.1.Definitions as Hypotheses

These definitions have content, but what do they do for researchers? Asking about the function of definitions reveals that there are questions that are not solely about whether a definition expresses the properties of a well delineated group of phenomena. Rather, hints at an alternative motivation are found in the writings of those who discuss memory regeneration and transfer. Baluška and Levin state, "the lessons we learn from unconventional cognitive systems will deeply impact our most basic understanding of how mind emerges from the brain" (2016, 12). Deochand and colleagues suggest that studies that demonstrate memory transfer "add valuable insight into our conceptualization of memory" (2018, 447). These quotes, though vague and hyperbolic, suggest that counting these discoveries as memory allows researchers to relate them to paradigmatic varieties of memory, reshaping how we understand memory in the process.

These hints illustrate what I call adopting *a definition as a hypothesis*: a contentful definition that picks out phenomena with its extension, which allows for the test of inferences. The content of a definition – for example, experience-dependent modification of internal

[3] For examples of other broad (or "extended") definitions of memory proposed by philosophers and scientists, see Zlotnik and Vansintjan 2019. Like the definitions I address, these definitions are disputed but play an unappreciated role in scientific inquiry.

structure that affects stimulus response – generates inquiry via testing similarities amongst phenomena. Like any hypothesis, these definitions are not formulated in a vacuum. Researchers draw from earlier findings when formulating them. Nonetheless, the content of the definition orients researchers to its test, and researchers adopt it because its content demarcates phenomena on which they test. They investigate phenomena to which the definition applies, which they may not do if it did not apply to these phenomena. Using findings on these phenomena as data, they evaluate whether the definition captures relevant similarities. Through these tests, researchers may determine additional similarities. Based on the results of these tests, they keep, reformulate, or abandon this definition. As a hypothesis, there is no assumption or commitment that there are relevant similarities; rather, the aim is to test for these similarities.

My account shifts our focus from solely what a definition expresses to a discussion about how definitions function in science. This shift is not without precedent. Feest claims that concepts serve as tools when they "enable experimental interventions in the process of investigating the purported or ill-understood object," even if they may not withstand the evaluation that results from this research (2010, 177). Likewise, Allen claims that defining 'cognition' involves "working definitions": definitions "suitable… for orienting newcomers to phenomena of potential interest," where "orienting towards phenomena worthy of further investigation does not depend on the kind of precisification that follows from detailed study of those phenomena" (2017, 4239). These claims highlight that definitions orient research instead of solely expressing research outcomes.

Bich and Green take a similar approach when defining 'life.' They claim that "rather than taking for granted that the role of definitions in science is equal to the demarcating natural kinds… we approach the issue of definitions from a pragmatic standpoint" (2018, 3921). Bich

and Green introduce what they call "operational definitions," which "combine, or integrate into a theoretical model, a set of mutually dependent necessary and satisficing criteria… that imply observable operations" (2018, 3933-3934). On their account, definitions "play an active operational role in guiding theoretical and experimental research" (Bich and Green 2018, 3933). Bich and Green address how definitions guide practice: in this case, deriving observable criteria from theoretical models.

These accounts explain how definitions pick out targets and guide researchers toward them, which is one reason why researchers call them 'definitions' and do not eliminate them from their lexicon. In thinking about definitions as *hypotheses*, I discuss an additional function that these definitions have. These definitions also can be used to test which properties of these targets cluster. The hypothesis can be construed as a question: if targets have the properties expressed in a definition, what (if any) other interesting properties cluster amongst them? Correspondingly, adopting a definition as a hypothesis consists in conjecturing about these clustering properties.

As with any hypothesis, definitions as hypotheses may be refuted. When is it fair to say that one of these hypotheses has been refuted? At minimum, there is a requirement of *novelty*. The hypothesis has been refuted if researchers do not uncover any similarities following testing that they did not have prior to testing. There is also a requirement of *non-superficiality*. I say more about superficiality in *Section 4*, but the key element is that similarities should inform researchers' ability to characterize, explain, or control their targets of investigation in line with their research aims. For instance, researchers may determine that the phenomena that count as

memory on a broad definition are "only superficially similar to one another" (Michaelian 2010, 170). If this knowledge is produced, then this definition should be abandoned.[4]

How does my view of definitions as hypotheses apply to memory? The researchers in *Section 2* formulated broad definitions based on earlier findings, including those from McConnell. They now test inferences about these definitions against the function of phenomena to which these definitions apply, allowing them to determine how these phenomena contribute to the activities of the systems in which they occur. For example, researchers now link memory regeneration with research on signaling processes that affect "embryogenesis, regeneration, and cancer suppression" (Levin and Martyniuk 2018, 82). Hypothesizing about a definition that applies to seemingly dissimilar signaling phenomena allows for "important and testable consequences" regarding the manipulation of memory and morphological states as well as how these states persist despite perturbation (Pezzulo and Levin 2015, 1495). Further, researchers link gene activations during these seemingly dissimilar phenomena in organisms ranging from aplysia to chimpanzees, which may elucidate genetic similarities (Levin and Martyniuk 2018, 88). While the results of these tests may not turn out to be fruitful, they nonetheless illustrate the strategy of adopting definitions as hypotheses.[5]

In addition, broad definitions generate inquiry about mechanistic schemata for phenomena, leading to tests of inferences about the schemata (or their components) that explain different phenomena. As exemplified by the research on naturalistic transfers of memory, researchers now investigate RNA interference regulation as a component of a mechanistic explanation for complex memory phenomena based on studies of sensitization and inheritance

[4] Clarifying the conditions under which a definition is refuted mitigates the chance that adopting a broad definition will "entrench our misconceptions" of memory (Cleland 2012, 128).
[5] The same is true about extended memory. Regardless of what one thinks about the validity of these definitions, they have generated scientific inquiry (Zlotnik and Vansintjan 2019).

phenomena (Mercer et al. 2008). Broad definitions also are used to test inferences about the mechanistic underpinnings of memory phenomena based on recently characterized epigenetic mechanisms (Chen and Schuman 2016) and non-neuronal bioelectrical patterning mechanisms (Baluška and Levin 2016, 4).

Like the broad definitions themselves, these tests are not limited to non-declarative memory. Researchers now compare mechanisms of phenomena that count as memory on a broad definition, such as the transplant or epigenetic inheritance examples, with those underlying hippocampal episodic memory. A significant debate in which this comparison contributes is in insights about the relative importance of molecules versus synapses in "memory engrams" (Josselyn, Köhler, and Frankland 2017; Asok et al. 2019; Josselyn and Tonegawa 2020; Langille and Gallistel 2020). Research informed by broad definitions not only supports a molecular memory mechanism with which researchers must now contend; it provides arrangements in which researchers can manipulate molecules or synapses and determine their respective effects on engram storage and organism behavior. These insights may have been missed if researchers had not adopted definitions that apply to transplants or epigenetics. Unlike definitions that only apply to episodic or even declarative memory, defining memory broadly extends the tradition of questioning established taxonomies based on mechanistic similarities amongst memory subtypes (Kandel 2009).

These examples show that broad definitions have generated inquiry in multiple areas of memory research, some of which extends all the way to complex, human-centric varieties like episodic memory. In addition, there is promise for further inquiry. Investigating potentially similar phenomena in seemingly dissimilar systems (some of which are in simple organisms) allows for novel intervention methods, which can inform tests about the clustering properties of

these mechanisms and the underpinnings of complex, human-centric varieties of memory. Again, these tests may not be fruitful, but the fact that there are tests supports my account. It is possible that these tests will not bear fruit in terms of general, novel, and relevant similarities between paradigmatic memory phenomena and phenomena that only count as a memory on a broad definition. I address this possibility in *Section 5*.

While my account explains the idea of a definition as a hypothesis, it also explains why, in the case of memory, the definitions are *broad* ones. Memory researchers aim to determine whether there are any interesting inferences to be made about targets that at first glance seem different: me remembering what I ate for breakfast and a planarian associating stimuli seem like very different phenomena. If researchers want to learn whether there are similarities between these phenomena, their definitions that orient them to test inferences must apply to all of their targets, and these definitions must be suitable for operationalization consistent with the disparate measures of these targets. The extension of a narrow definition of memory that appeals to representation or phenomenology does not cover this wide a range of phenomena: a rich, narrow definition that expresses properties of declarative memory does not apply to phenomena in organisms like planaria and aplysia, nor can it be operationalized for their measurements.[6] Hence, we find definitions that apply to paradigmatic memory phenomena and controversial phenomena.

This advantage of broad definitions is supported by the fact that defining 'memory' broadly is not a new idea in memory research. Squire notes that "neuroscientists and psychologists have generally preferred broad definitions of learning and memory," as "a complete account of learning and memory must accommodate many different kinds of

[6] This is not to say that a narrow definition adopted as a hypothesis can never be fruitful. The definition need only allow for the test of inferences.

behavioral change" (1987, 3). This appeal to broad definitions has been driven by memory researchers' perceived limitations in their knowledge about their targets: "at present [the 1980s] we understand them so poorly that we cannot afford to discard any phenomenon that might yield clues about their neural basis" (Squire 1987, 5).

The reader may be sympathetic to definitions as hypotheses but remain skeptical that, in this case, this definition should be *of memory* rather than of a distinct category all together. To draw an analogy, I imagine that most philosophers would agree that the fact that there are similarities amongst mammals does not warrant a definition of 'human' that counts all mammals as humans, even if this definition is proposed as a hypothesis. How do my claims about memory differ? First, the definitions that I discuss are not best conceived of as a radical revision to our current scientific (Dudai 2007), philosophical (De Brigard 2014), or folk understanding of memory. Neither definition denies that paradigmatic cases of memory, such as me remembering what I ate for breakfast, count as memories. Second, these broad definitions of memory do not supplant definitions for a well-specified and supported category that researchers investigate. In the case of memory, there remains a debate over whether there is a scientifically valuable definition of 'memory' to be devised and, if so, what this definition should be. This is unlike the mammal case: one reason to not define 'human' broadly is that it is useless if it picks out the same category that 'mammal' picks out.

3.2. Definitions and Kinds

A definition's extension picks out a kind. By 'picks out,' I mean that the definition's extension circumscribes a kind whose members' properties are expressed by the definition. Adopting a definition as a hypothesis allows researchers to determine what properties cluster

amongst members of the kind it picks out, and, correspondingly, they can determine what claims about this kind's members project.

My position commits me to a distinction between characterizing a phenomenon's properties versus counting this phenomenon as a memory phenomenon. The former is a descriptive project: determining that a phenomenon with certain characteristics occurs and formulating these characteristics into a representation (Colaço 2020). The latter involves determining whether the characteristics of the phenomenon warrants categorizing it with other targets of research. In practice, however, these two projects can intertwine. The historical research in *Section 2* illustrates that researchers often characterize a phenomenon in terms of what kind of phenomenon they think it is. Memory researchers of the 60s thought of memory chiefly in associative terms, which informed McConnell's experimental tests. When he believed that he obtained association transfers, he concluded that he had characterized a novel memory phenomenon. Nevertheless, the two projects can be independently evaluated: a phenomenon's characterization can be adjudicated independently of whether it counts as a memory phenomenon.

My account of definitions as hypotheses relates to what Brigandt calls an "investigative kind," or "a group of things that are presumed to belong together due to some underlying mechanism or a structural property," where "the idea that these entities belong to a kind might be due to some interesting similarities" (2003, 1309). I am sympathetic to the idea that "an investigative-kind concept goes together with a scientific search, which might be open-ended" (Brigandt 2003, 1310), where these "concepts are ongoing projects of inquiry" (Griffiths 2004, 907). Brigandt is right to point out an epistemological fact about scientific kinds: these kinds are often conceptualized during on-going research. This fact thematically underwrites my account.

However, the difference between my account and Brigandt's is that, on my account, it need not be the case that the kind picked out by a definition adopted as a hypothesis "is specified by some nontrivial underlying feature or process that is presumed to account for the observed similarities," as he argues for an investigative kind-concept (Brigandt 2003, 1309). This is because, as I show in the case of memory, neither the similarities nor the explanations for these similarities is well-understood. Rather, determining whether there are interesting similarities and an explanation for them are both aims of the research facilitated by a definition as a hypothesis. Thus, researchers may know even less about their targets when adopting a definition as a hypothesis than they know when they adopt investigative kind-concepts.

If there is no specification of a "nontrivial underlying feature or process" to support the adoption of a definition as a hypothesis, what separates appropriate cases of this adoption from inappropriate cases? The answer to this question lies in the fact that the definition picks out a category of phenomena that can all be described in a consistent and contentful manner. In the case of memory, a broad definition extends to paradigmatic cases of memory (such as me remembering what I ate for breakfast), and it extends to controversial phenomena and phenomena whose properties are still under investigation. This results in a kind for which the projectability of claims about all of these phenomena can be adjudicated. Thus, the success of a definition as a hypothesis depends on the discovery and characterization of phenomena that broadly count as memory, and this success is evaluated via attempts to project claims about the members of the kinds that are picked out. If the researchers who adopt a definition as a hypothesis do not test it, then the definition ceases to be scientifically valuable and therefore should be abandoned.

This account lays out the foundations for what I call *kinding in progress*. Here, I understand "kinding" as the practice of discovering useful categories that can be considered to be valuable kinds (Kendig 2016, 107). I use the phrase 'kinding in progress' for the categorization of targets of investigation as members of a kind when the process to discover and determine the properties of the targets to be categorized is on-going, and, therefore, many of these properties are unknown at the moment of the categorizing in question. In the case of memory, kinding in progress is facilitated by a definition as a hypothesis, as this definition orients researchers towards the test of inferences. While this strategy can fail for a variety of reasons, it is not my aim to defend it as the ideal strategy in memory research. Rather, I explain why researchers might choose to propose broad definitions of memory that rebuts the negative responses I presented in the introduction. Let's look at these negative responses, as they motivate potential objections to my account.

4. The Dissimilarity Objection

The responses in the introduction raise the objection that a broad definition of memory lumps together dissimilar phenomena. Whether the definitions pick out phenomena that "fall outside the rubric of memory" (Schacter 2007, 24), or they extend to "phenomena only superficially similar to one another" (Michaelian 2010, 170), these objections converge on the idea that there are relevant dissimilarities between paradigmatic memory phenomena and the phenomena that count as memory only on a broad definition. A skeptic of these broad definitions might reject the idea that these definitions support the right inferences about memory. There are, the skeptic might argue, too many differences between (say) episodic memory in humans and

long-term sensitization in aplysia to count them as anything but superficially similar. If correct, this objection undermines the scientific value of a broad definition of memory.

What are the similarities that a skeptic might think are relevant? One set of criteria come from Michaelian, who argues that declarative and nondeclarative memory are dissimilar because they are individuated by different memory systems, and these systems do not share computational-, algorithmic-, or implementational-level descriptions (2010, 175). According to Michaelian, they "do not perform a common information-processing task," as "we need not mention information-processing at all to give a complete description of nondeclarative memory"; further, their individuating systems are localized to different neural structures (2010, 178). Thus, according to Michaelian, conceiving of memory as a singular kind results in lumping together dissimilar phenomena if the individuating systems of these phenomena cannot be commonly described at the three Marrian levels of analysis. If one accepts Michaelian's claims, then long-term sensitization (or the transfer thereof) is not relevantly similar to episodic memory if a shared information processing description does not elucidate this sensitization phenomenon.[7]

Does the argument that Michaelian presents serve as a good reason against adopting a broad definition of memory as a hypothesis? I say no for three reasons. First, it is not obvious that the properties of newly discovered phenomena cannot be fruitfully described in terms of information processing. Information processing constructs have been used by some proponents of broad definitions of memory. Baluška and Levin claim that "the coverage of cognitive terms [like memory] across biology can expand to the extent that information-centered approaches are shown to be effective in predicting and controlling the behavior of biological systems,"

[7] Michaelian is not the only one who suggests that information processing does not contribute to describing non-declarative memories. Squire claims that non-associative sensitization does "not constitute [a] memory system as much as different reflex pathways," suggesting that differences between memory systems is so great that a definition of memory that extends to all of them may not be valuable (2007, 342).

suggesting that there may be value to conceiving of novel phenomena in terms of information processing (2016, 12). While this might reflect a "bandwagon" application of information terminology (Shannon 1956), this alone is not a reason to rule out conceiving of phenomena that occur in "unconventional cognitive systems" in terms of information processing.

Second, as the discovery of phenomena is on-going, much about these novel phenomena is simply not known. As a result, it is unclear which similarities will be discovered, let alone whether these similarities are superficial. *Section 3.1.* illustrates insights about memory engrams based on comparing newly discovered "non-declarative" transplant and epigenetic phenomena with episodic memory. This example shows that, if the similarities are not yet known, we are not able to adjudicate the commonality of descriptions of the function or implementation of the systems in which these newly discovered phenomena occur. Using these definitions to determine the similarity of phenomena is precisely why a definition as a hypothesis can be scientifically valuable in on-going research. And, before this research is performed, we are not in a position to determine that dissimilar phenomena are lumped together if one adopts a broad definition of memory.

Third, determining the dissimilarity of memory phenomena based on their Marrian descriptions is only one approach to arguing that a broad definition lumps together dissimilar phenomena, which Michaelian admits (2010, 175). Alternative criteria for assessing whether phenomena are similar can be devised, and these alternative criteria will invariably reflect the aims of the researchers in question. These aims, such as characterizing, explaining, and controlling phenomena, may be achieved by focusing on different similarities or differences between these phenomena. Hochstein defends the role of aims in categorizing mental phenomena, noting that "the assumption that there is a single ideally correct way of classifying

or categorizing mental phenomena… runs counter to productive scientific practices in these domains" (2016, 746). Researchers study phenomena in order to achieve aims and goals, and the "best classificatory scheme to use for some of these goals will not be the same as the one best suited for others" (Hochstein 2016, 746).

Thus, it is not only the case that researchers are not yet in a position for them to determine if newly discovered phenomena are similar to paradigmatic cases of memory. It is also that case that the relevance of these similarities cannot be adjudicated without understanding the aims of the researchers. This consideration of aims does not entail that "anything goes" when it comes to researchers' aims, nor does it entail that understanding researchers' aims is straightforward. What matters for my response to the dissimilarity objection is that what guides researchers' approach to categorization influences how they adjudicate the relevance of the similarities amongst their targets. This fact provides reason to resist the idea that the phenomena "lumped" together on a broad definition of memory are superficially similar.

5. The Natural Kinds Objection

The second objection I address is that a definition of memory does not pick out a natural kind, and this failure indicates its bankruptcy as a valuable scientific definition. This is stronger than the dissimilarity objection. It is not merely the case that a *broad* definition of memory is deficient: any definition of memory as a superordinate kind is deficient if memory is not a natural kind. It remains a popular idea that natural kinds support inferences, regardless of if one is a neo-essentialist (Putnam 1973), property cluster theorist (Boyd 1991), or causal node theorist (Khalidi 2018). All of these accounts capture that a natural kind "had better attach not just to an accidental pattern of correlated properties, but to proper ties correlated for a good reason"

(Millikan 2000, 17). Machery puts the position this way: "a natural kind is a class about which many generalizations can be formulated [, and] these generalizations are not accidental" (2009, 232). While some philosophers suggest that natural kinds can be accounted for in terms of the "stability a cluster of properties might possess" (Slater 2015, 396), most argue that "there is at least one causal mechanism that explains why its members tend to have those properties" (Machery 2009, 233). If the members of the generic memory kind do not share properties and a corresponding explanation for these properties, then generic memory is not a natural kind. Showing that memory is not a natural kind may warrant eliminativism, akin to what has been argued for "concept" (Machery 2009). This is the ultimate conclusion that Michaelian (2010) draws via the dissimilarity objection.[8]

For what reason might one argue that generic memory is not a natural kind? Rupert argues that "the positing of a weakly defined type, generic memory, does not improve prospects… [because] there is little, if any, causal-explanatory work for such a watered-down kind to do" (2004, 407). Thus, Rupert's objection to accounts that appeal to "generic memory" is that there is no shared explanation for all memory subtypes and positing generic memory is explanatorily empty.[9] Does this argument provide reason to reject a broad definition of (generic) memory as a hypothesis?

I say no for two reasons. First, I have provided support for the idea that researchers are not in a position to adjudicate the commonality of explanations amongst paradigmatic memory phenomena and newly discovered phenomena. Those who propose the broad definitions of

[8] While Michaelian claims that memory is not a natural kind, he accepts that subtypes of memory are natural kinds. Tulving comes to a similar conclusion when he claims that "no profound generalizations can be made about memory as a whole," suggesting that a definition of "memory as a whole" is not valuable (1985, 385).
[9] Rupert's criticisms originally were directed at extended memory. For a review of responses to Rupert, see Menary 2010.

memory that I introduced "motivate such a broad definition of 'memory' by putting the definition to work… by running new experiments to demonstrate [its] value" (Rupert 2004, 409). This is illustrated by the tests that these researchers perform. This, I argue, is the function of a definition of memory as a hypothesis, which provides a rebuttal to Rupert's claim that it "is not clear what useful role the kind generic memory plays in any real-life research program" (2004, 420).

Second, Rupert's argument focuses on the putative lack of "explanatory power" of generic memory, which, I argue, is the wrong virtue to care about when defining generic memory as a hypothesis (2004, 390). Rupert's focus is captured in his question: "what data is the introduction of 'memory'… meant to help us to systematize or explain?" (Rupert 2013, 33). In response to Rupert's question, the broad definitions of memory that I have introduced play a systematizing role: they circumscribe phenomena against which researchers test these definitions. Endorsing a definition of generic memory as a hypothesis provides no explanation of the properties of memory. However, this is not a deficit of my account. Explanation is not the aim of adopting a broad definition of memory as a hypothesis. Rather, it is the goal of adopting this definition. The virtue of broad definitions of memory as hypotheses is their facilitation of testing inferences about phenomena. Thus, Rupert's criticisms do not serve as good reasons to reject definitions of (generic) memory as hypotheses.

My responses to Rupert highlight the epistemic dimension of "kinding in progress." My account of a definition as a hypothesis implies that researchers should adopt these definitions if they do not know whether a domain of phenomena have properties that support inferences. Therefore, it is possible that these definitions of memory do not pick out natural kinds. If this is the case, what should we conclude about broad definitions of memory? Here, it is important to take account of the present research context. That memory phenomena do not share the right

properties or share them by accident is a fact that may not be known until researchers investigate these phenomena. One way to determine that memory is not a natural kind is via the strategy that I have introduced: adopt the definition as a hypothesis and see what inferences are supported when the members' properties are better understood. Following this test, either a better definition memory can be adopted, or we can eliminate generic memory and "replac[e] it with more appropriate theoretical terms" (Machery 2009, 230). Thus, even if broad definitions do not pick out natural kinds, they are valuable because they help researchers to determine that they do not pick out natural kinds.

The epistemic uncertainty of "kinding in progress" should lead philosophers to show humility with respect to what properties might be discovered to be similar amongst memory phenomena. Due in no small part to the intuitions generated from examples like 'water is H_2O,' the reader might think that the properties that unite any natural kind are "neat and tidy" when they are discovered (Havstad 2018, 720). However, this view about kinds is less plausible when one takes into account claims that "chemical kinds such as elements (like gold) and compounds (like water) are themselves rather messy," and "the foregoing philosophical picture of chemical kindhood has been based on an extremely tenuous inference from an oversimplified discussion of these sparse paradigm cases" (Havstad 2018, 720). In this context, 'messy' signifies that complications and disjunctive claims must be added to any description of a kind for it to capture this kind's microstructural and macrostructural properties (Havstad 2018, 725).

If a detailed analysis of chemistry reveals that chemical kinds, the benchmark examples of natural kinds, are actually "rather messy" but nevertheless valuable to chemists, then we should be open to the idea that cognitive kinds like memory are messy as well. This conclusion cuts both ways. It will likely be the case that, if memory is a natural kind, it will be best

described with far more complexity than the broad definitions of memory express. However, drawing on the clustering properties of function, genetics, and mechanisms I discuss in *Section 3.1.*, it is also plausible that the properties that researchers will end up picking out as clustering amongst memory phenomena will also be complex, and these properties may not yet have been identified let alone understood. If the definitions generate inquiry that results in these sorts of discoveries, then these definitions' outcomes may be reformulation or abandonment. These outcomes are fine, as the definitions did the job for which they were adopted.

Nonetheless, my invocation of "messy kinds" highlights the importance of engaging in kind discovering practices and being open to the diverse outcomes of these practices. Certainly, from "where we have already arrived," we do not have any reason to presuppose that the kind picked out by a broad definition supports the right inferences to count as a natural kind (Rupert 2013, 38). However, this conclusion does not undermine my account of broad definitions of memory as hypotheses, as the value of a definition as a hypothesis does not rest on whether it picks out a natural kind.

6. The Knowledge Objection

An ardent reader of the philosophical and scientific literature on memory might be perplexed by my suggestion that definitions of memory as hypotheses are scientifically valuable because the shared properties of memory phenomena are largely unknown. After all, even on a very conservative estimate, the rigorous study of memory has been in practice for well over a century. While it may have been arguable that we understood the targets of memory research poorly in the late 1980's (Squire 1987), distinct kinds of memory have now been known for decades (Squire 2004). Likewise, some philosophers suggest that we know enough about the

neural basis of memory phenomena to inform our taxonomy of memory (Bickle 2003). Why would researchers need to "kind in progress" at this time?

An immediate response to the knowledge objection is that there are new phenomena being discovered, and, because of these discoveries, there is reason to think that using definitions as hypotheses to engage in "kinding in progress" makes sense. However, this immediate response is unsatisfactory because it begs the question at hand. The skeptic might not deny that there are discoveries putatively being made, but this is not the relevant question. The question, the skeptic might continue, is whether these discoveries are pertinent to defining 'memory,' so we should not have to assume that these discoveries are pertinent to memory research in order to motivate broad definitions of memory as hypotheses.

I propose a more nuanced response to the knowledge objection. To begin, it is important to be clear about what is known in contemporary memory research. Given that I distinguish between characterizing phenomena and determining whether they count as memory phenomena, I do not deny that researchers have made significant advancements in characterizing phenomena and explaining them. These advancements need not be undermined if one adopts a definition of memory as a hypothesis. While the characterization of phenomena is ongoing, it need not be the case that this knowledge will require us to revise what we think (say) episodic memory is.

However, the same cannot be said about counting a phenomenon as a memory phenomenon. Often times in cognitive science, researchers find themselves in a situation where they discover phenomena that they want to relate to phenomena that they already have characterized. The question is how they should they go about doing this. Relying solely on an understanding of the relations between known phenomena is insufficient, as the introduction of new potential relata opens up the possibility for the discovery of new sorts of relations that may

be valuable to scientists. Why think that there might be these relations? To answer this, I return to the idea that both broad definitions of memory that I have discussed throughout this paper have content. It is the case that both paradigmatic cases of memory and newly discovered phenomena are similar to one another in the sense expressed in these definitions, and, given the content expressed, this similarity is not vacuous. While this similarity may turn out to be without scientific value, there is no reason to presuppose this outcome. The targets share properties, so there is a valid question to be asked about what other properties they may share. Instead of presupposing anything about their relations based on previous evidence, researchers instead can test these relations in light of new discoveries.

The reason that the knowledge objection does not undermine the adoption of a broad definition of memory as a hypothesis is that, even if we already know a lot about the properties of different memory phenomena, there remain factors that we can learn about memory and the inferences that our conceptualization of memory can support. We need not deny that the distinctions made in (say) Squire's (2004) taxonomy are empirically supported when we question whether this taxonomy and the kinds contained therein are exhaustive of the different memory phenomena about which we can generalize or extrapolate. The categorization of phenomena as memory phenomena – and, correspondingly, the inferences about these phenomena that are supported – is a research project that should be engaged in when phenomena are discovered and characterized that have similarities to known phenomena. New targets to "kind" can require a revision to our concepts, especially when the properties of these new targets offer an avenue to investigate known phenomena in more depth and detail. In this way Squire's (1987) concern is still relevant to the current state of memory research: when new phenomena are discovered that are similar in respects to paradigmatic memory, it is reasonable for scientists to not discard any

of these phenomena "that might yield clues" about the properties of memory that support inferences, which inferences are supported, and what their value to scientists might be.

There is a final point to be made regarding definitions of memory in the long-term trajectory of the field. A skeptic of broad definitions of memory might identify the disagreement amongst scientists and philosophers who aim to define memory as an "embarrassment of riches." To compare the case to debates about defining 'life,' Machery claims that quibbles over the correct scientific definition are pointless because these disciplines have different aims and seek to tailor their scientific definitions to different targets (2012, 158). This sort of "embarrassment of riches" worry casts doubt on the idea that memory researchers ever will develop a univocal definition of memory that is valuable to all researchers and philosophers. One definition of memory probably cannot do all the jobs that philosophers and scientists want a definition of memory to do, and, because of this, any fight over which definition is the best definition of memory is unwinnable and unproductive.

At no point in this paper have I argued that a broad definition of memory should be a univocal definition of memory. In fact, my discussion of multiple definitions and my emphasis on the distinct functions of definitions highlights that we must ask not only what properties a definition of memory expresses but also for what function these properties are expressed. Further, we must ask whether the properties expressed in the definition of memory help researchers fulfill the aims that they have for proposing this definition. These functions capture different aims of the researchers, so there is no reason to assume that the properties expressed in these definitions will be univocal. Definitions of memory as hypotheses is one way of defining 'memory' that is proposed to fulfill an aim that the researchers mentioned in *Section 2* have: they aim to determine whether inferences about phenomena that count as memories on one of these

definitions are supported. Because the discovery of these phenomena is on-going, this situation is not best described as an "embarrassment of riches": there remains empirical and conceptual work to be done, which can influence how we answer the question of what counts as a memory.

7. Conclusion

I have explained why researchers might adopt what I have called 'broad definitions of memory,' and I have explained the scientific value of investigating the kinds that are picked out by these definitions. According to my account, these definitions are proposed as hypotheses to test inferences against memory phenomena (so defined). My account is not struck down by the idea that these definitions lump together dissimilar phenomena, nor is it undermined by the idea that generic memory is not a natural kind. While it is no doubt the case that several practical factors, such as historical uses of the term 'memory,' prestige, and access to funding, also motivate researchers to construe their experimental findings in terms of 'memory,' there remains good scientific reason to broadly define memory.

My position points to the idea that the question "what counts as a memory?" has at least two different senses. The first sense can be construed as asking for an expression of the criteria for counting a given target as a memory based on what is already known about this and other targets. The second sense can be construed as asking for an expression of the criteria researchers should use to pick out targets that will facilitate their investigation of these targets. These are different questions, and the latter has received insufficient attention in the philosophical literature. This fact, compounded with research on newly discovered phenomena, highlights the need to address how definitions inform rather than solely result from empirical investigations.

Adopting a broad definition of memory as a hypothesis lives or dies on the ability of researchers to successfully discover and identify the properties of phenomena that fit the content of this definition, so that they can determine their similarities, their relevance, and their cause. In this paper, I have only provided limited discussion of what counts as discovering a target phenomenon. Nonetheless, it is from the perspective of the context of discovery that this sort of 'kinding in progress' is best understood. Researchers develop definitions to pick out kinds not to express similarities and to support of inferences about memory but rather to determine similarities and the potential for inferences about memory.

References

Allen, Colin. 2017. “On (not) defining cognition.” *Synthese*, *194*(11), 4233-4249.

Anderson, John Robert. 1990. *The adaptive character of thought*. Hillsdale, NJ: Lawrence Erlbaum.

Asok, Arun, Félix Leroy, Joseph B. Rayman, and Eric R. Kandel. 2019. "Molecular mechanisms of the memory trace." *Trends in neurosciences* 42, no. 1: 14-22.

Baluška, František, and Michael Levin. 2016. “On having no head: Cognition throughout biological systems.” *Frontiers in psychology*, *7*, 902.

Bédécarrats, Alexis, Shanping Chen, Kaycey Pearce, Diancai Cai, and David L. Glanzman. 2018. "RNA from trained Aplysia can induce an epigenetic engram for long-term sensitization in untrained Aplysia." *eNeuro*, *5*(3).

Bich, Leonardo, and Sara Green. 2018. “Is defining life pointless? Operational definitions at the frontiers of biology.” *Synthese*, *195*(9), 3919-3946.

Bickle, John. 2003. *Philosophy and neuroscience: A ruthlessly reductive account*. Dordrecht, The Netherlands: Kluwer Academic.

Brigandt, Ingo. 2003. “Species pluralism does not imply species eliminativism.” *Philosophy of Science*, *70*(5), 1305-1316.

Boyd, Richard. 1991. “Realism, anti-foundationalism and the enthusiasm for natural kinds.” *Philosophical studies*, *61*(1), 127-148.

Carney, Rosalind. 2018. “An emerging role for RNA in a memory-like behavioral effect in Aplysia.” *eNeuro*, *5*(3).

Chen, Wei, and Erin Schuman. 2016. "Circular RNAs in brain and other tissues: a functional enigma." *Trends in Neurosciences*39, 9, 597-604.

Cleland, Carol. 2012. “Life without definitions.” *Synthese*, *185*(1), 125-144.

Colaço, David. 2018. “Rip it up and start again: The rejection of a characterization of a phenomenon.” *Studies in History and Philosophy of Science Part A*, *72*, 32-40.

Colaço, David. 2020. “Recharacterizing scientific phenomena.” *European Journal for Philosophy of Science*, *10*(2), 1-19.

De Brigard, Felipe. 2014. “The nature of memory traces.” *Philosophy Compass*, *9*(6), 402-414.

Deochand, Neil, Mack Costello, and Michelle Deochand. 2018. "Behavioral research with planaria." *Perspectives on Behavior Science*, *41*(2), 447-464.

Dudai, Yadin. 2007. "Memory: It's all about representations." *Science of memory: Concepts*, 13-16.

Feest, Uljana. 2010. "Concepts as tools in the experimental generation of knowledge in cognitive neuropsychology." *Spontaneous Generations: A Journal for the History and Philosophy of Science*, *4*(1), 173-190.

Figdor, Carrie. 2018. *Pieces of mind: The proper domain of psychological predicates*. Oxford University Press.

Griffiths, Pail. 2004. "Emotions as natural and normative kinds." *Philosophy of science*, *71*(5), 901-911.

Havstad, Joyce. 2018. "Messy chemical kinds." *The British Journal for the Philosophy of Science*, *69*(3), 719-743.

Hochstein, Eric. 2016. "Categorizing the mental." *The Philosophical Quarterly*, *66*(265), 745-759.

Josselyn, Sheena A., Stefan Köhler, and Paul W. Frankland. 2017. "Heroes of the engram." *Journal of Neuroscience* 37, no. 18: 4647-4657.

Josselyn, Sheena A., and Susumu Tonegawa. 2020. "Memory engrams: Recalling the past and imagining the future." *Science* 367, no. 6473.

Kandel, Eric R. 2009. "The biology of memory: a forty-year perspective." *Journal of Neuroscience* 29, no. 41: 12748-12756.

Kendig, Catherine. (Ed.). 2015. *Natural kinds and classification in scientific practice*. Routledge.

Khalidi, Muhammad Ali. 2018. "Natural kinds as nodes in causal networks." *Synthese*, *195*(4), 1379-1396.

Langille, Jesse J., and Charles R. Gallistel. 2020. "Locating the engram: Should we look for plastic synapses or information-storing molecules?." *Neurobiology of Learning and Memory* 169: 107164.

Levin, Michael, and Christopher J. Martyniuk. 2018. "The bioelectric code: An ancient computational medium for dynamic control of growth and form." *Biosystems* 164: 76-93.

Machery, Edouard. 2009. *Doing without concepts*. Oxford University Press.

Machery, Edouard. 2012. “Why I stopped worrying about the definition of life... and why you should as well.” *Synthese*, *185*(1), 145-164.

McConnell, James. 1962. “Memory transfer via cannibalism in planaria.” *J. Neuropsychiat*, *3*, 1-42.

McConnell, James, Allan Jacobson, and Daniel Kimble. 1959. “The effects of regeneration upon retention of a conditioned response in the planarian.” *Journal of comparative and physiological psychology*, *52*(1), 1.

Menary, Richard. 2010. "Introduction to the special issue on 4E cognition." *Phenomenology and the Cognitive Sciences*, *9*(4), 459-463.

Mercer, Tim R., Marcel E. Dinger, Jean Mariani, Kenneth S. Kosik, Mark F. Mehler, and John S. Mattick. 2018. "Noncoding RNAs in long-term memory formation." *The Neuroscientist*, *14*(5), 434-445.

Michaelian, Kourken. 2010. “Is memory a natural kind?.” *Memory Studies*, *4*(2), 170-189.

Millikan, Ruth. 2000. *On clear and confused ideas: An essay about substance concepts*. Cambridge University Press.

Moore, Rebecca, Rachel Kaletsky, and Coleen Murphy. 2018. “C. elegans pathogenic learning confers multigenerational pathogen avoidance.” *bioRxiv*, 476416.

Pezzulo, Giovanni, and Michael Levin. 2015. "Re-membering the body: applications of computational neuroscience to the top-down control of regeneration of limbs and other complex organs." *Integrative Biology* 7, no. 12: 1487-1517.

Posner, Rachel, Itai Antoine Toker, Olga Antonova, Ekaterina Star, Sarit Anava, Eran Azmon, Michael Hendricks, Shahar Bracha, Hila Gingold, and Oded Rechavi. 2019. "Neuronal small RNAs control behavior transgenerationally." *Cell* 177(7): 1814-1826.

Putnam, Hillary. 1973. “Meaning and reference.” *The journal of philosophy*, *70*(19), 699-711.

Rupert, Robert. 2004. “Challenges to the hypothesis of extended cognition.” *The Journal of philosophy*, *101*(8), 389-428.

Rupert, Robert. 2013. “Memory, natural kinds, and cognitive extension; or, Martians don’t remember, and cognitive science is not about cognition.” *Review of Philosophy and Psychology*, *4*(1), 25-47.

Schacter, Daniel. 2007. “Memory: Delineating the core.” *Science of memory: Concepts*.

Shannon, Claude. 1956. “The bandwagon.” *IRE Transactions on Information Theory*, *2*(1), 3.

Shomrat, Tal, and Michael Levin. 2013. “An automated training paradigm reveals long-term memory in planarians and its persistence through head regeneration.” *Journal of Experimental Biology*, *216*(20), 3799-3810.

Slater, Matthew. 2014. “Natural kindness.” *The British Journal for the Philosophy of Science*, *66*(2), 375-411.

Squire, Larry. 1987. *Memory and brain.* Oxford University Press.

Squire, Larry. 2004. “Memory systems of the brain: A brief history and current perspective.” *Neurobiology of learning and memory*, *82*(3), 171-177.

Squire, Larry. 2007. “Memory systems: A biological concept.” *Science of memory: Concepts*, 339-343.

Thompson, Robert, and James McConnell. 1955. “Classical conditioning in the planarian, Dugesia dorotocephala.” *Journal of comparative and physiological psychology*, *48*(1), 65.

Tulving, Endel. 1985. “How many memory systems are there?.” *American psychologist*, *40*(4), 385.

Zlotnik, Gregorio, and Aaron Vansintjan. 2019. "Memory: An extended definition." *Frontiers in Psychology* 10 (2019).