

ORIGINAL ARTICLE

Memory is a modeling system

Sara Aronowitz

Program in Cognitive Science, Peretsman Scully Hall, Princeton University, Princeton, New Jersey

Correspondence

Sara Aronowitz, Program in Cognitive Science, Peretsman Scully Hall, Princeton University, Princeton NJ 08540.
Email: aronowitz@princeton.edu

This paper aims to reconfigure the place of memory in epistemology. I start by rethinking the problem that memory systems solve; rather than merely functioning to store information, I argue that the core function of any memory system is to support accurate and relevant retrieval. This way of specifying the function of memory has consequences for which structures and mechanisms make up a memory system. In brief, memory systems are modeling systems. This means that they generate, update and manage a series of overlapping, simplified, relational representations that map out features of the world. Succeeding at building and maintaining models requires the kind of active knowledge generation traditionally associated only with deliberative reasoning.

KEYWORDS

coherence, memory consolidation, models, rationality over time, representation, sleep

1 | INTRODUCTION

Let's start with an analogy. Some philosophers think that memory works like a museum. It is a place where old things are collected and stored over time. These old things are important not just because they are interesting and informative, but also because they are authentic links to our shared past. The function of the museum is simple: be a receptacle for these old things. A good museum is the right temperature so that the artifacts do not degrade—and likewise, on this picture, a good memory system maintains a cognitive environment to keep the items stored in memory in as close to their original condition as possible. In any real museum, items do degrade and need to be touched up and patched. But the idea of keeping things in the original condition is a guiding norm. Similarly, while human memory might involve alterations to the stored information, the system is still aiming at or guided by the norm of perfect preservation. A recent trend, constructivism, rejects the museum picture entirely. For instance, Michaelian (2016) makes the case that memory just *is* a kind of imagining—it has no special relationship to preserved content at all.

On the view I develop in this paper, memory is indeed analogous to a museum, though not in the way I have just described. To explain, I will sketch a different picture of how museums function. While they do usually contain some old and authentic artifacts, there is more to a good museum than good preservation. For one, there is a team of curators who arrange exhibits. This involves choosing what to display—but perhaps more importantly, organizing a series of objects around a theme or narrative. For instance, a curator might plan an exhibit about Turgenev. She would choose an angle—say, exploring his literary circle and emphasizing connections between France, Russia and England. The choice of a theme will depend on the artifacts at her disposal, but it also will determine which she chooses to display. The same goes for the process of arranging the exhibit and generating displays like maps, captions and installations. In the process of making a map, she will realize it would be helpful to display a postcard from Turgenev's travel to Italy or in the process of arranging a set of letters, she will realize it would be helpful to have a chart of his correspondence. A museum, in this sense, is an active institution as opposed to a passive repository. And this is most obvious when we consider the distinctive job of curation that occurs in museums.

Thus on this analogy, we can position alternative views of the computational function of memory on a scale from memory having the most active role in information operations to the most passive role. On the first extreme, the exhibits in the museum might be purely generated by the curators, paying no special attention to received artifacts and relying on minimal external resources. These “curators” would do nothing distinctively historical or geared towards long-term preservation. This view is of course memory constructivism. On the other extreme, which is the more traditional memory preservationism, the curators would merely carry out instructions given to them from outside. Note that at either extreme, curation is no longer particularly interesting—in the former case because it is not at all distinctive and in the latter because it is not at all active.

In this paper, I will argue that memory is somewhere between these two extremes, and operates a lot like an actual museum. By this, I mean that the memory system performs a distinctive epistemic activity which is analogous to curation. This operation, which I will claim is a kind of modeling, is aimed at structuring and altering stored contents in order to make correct, useful and relevant information available for retrieval. I will contrast this picture to both the initial museum analogy, where memory is aimed at authenticity and preservation, and to the constructivist account on which memory is a system of imaginative simulation. Just as curatorial work is mainly done when the museum is closed to visitors, the memory operations I am interested in occur during sleep and in the background during waking, rather than being conscious, on-line processes.

This paper aims to answer two questions in the epistemology of memory: (a) what is the epistemic role of memory, that is, the role of memory in generating knowledge? and (b) which features of the memory system are epistemically evaluable? My answer to these normative questions is inspired by recent work in computer science and neuroscience that has allowed us to dissect natural and artificial memory systems and better understand the problems which they succeed at solving.

While this account is about memory in general, I make several assumptions about the evidential context of the agent in question: (a) the agent has a set of evidence which is small relative to the total possible evidence about propositions which interest the agent, (b) the agent can expect to get more information over time that (at least in part) bears on the same propositions and (c) the agent is handling a large amount of complex information. Roughly, human memory falls into a class of systems solving a common problem: how to use large quantities of information under conditions of uncertainty when you expect to accumulate more and more in the future. This could apply to humans or other agents such as rats or primates, as well as machines that fall short of full agency, such as a database or specialized artificial intelligence.

Conditions (a)–(c) are plausible claims about most of our cognitive circumstances, but they are meant to capture something more specific: namely, what it means to be in the beginning of inquiry. I will argue that given these conditions, memory is not just a system for storing information, but actually has its own procedures and norms for generating new content and revising old content. On this picture, memory is differentiated from deliberation only by its degree of holistic connection and overall conservativeness.

The structure of this paper is as follows: in Section 2, I lay out the claim that retrieval is central to memory, and support it with empirical evidence that suggests retrieval is the most significant bottleneck in successful memory functioning. The museum analogy suggests that our memory systems are doing work in the background to update and change stored information; in Section 3, I show that active information transformation is happening during sleep in humans and other animals based on neuroscientific research. Understanding which transformations are happening and why requires a conceptual model. In Section 4, I describe two informational structures that solve the retrieval problem—first, a model system that structures the data in a series of simplified models, and second, an index system, which uses an intermediate structured representation between the query and the underlying unstructured storage system. I define these two types and present instances of real-world model and index devices. In Section 4.3, I argue that the model system is a better fit for memory in an ideal agent, by showing that as computational constraints are relaxed, index systems transform into model systems. In Section 5, I survey some views of memory in the philosophical literature, and show that they are problematic given the norm described in Section 2. I conclude by considering an objection that the model view is not at the right level of abstraction, and discussing some consequences of the model view.

2 | FROM STORAGE TO RETRIEVAL

Here is a seemingly trivial fact: memory systems must make their stored information available for retrieval. However, this fact has been given short shrift in the current epistemological literature. If we think of memory as a process that begins with some experience and leads to a later experience of remembering, most philosophers have focused on the experience of remembering itself. Memory constructivists like Michaelian (2016), for instance, are interested in how we are able to fill out the details and construct a model of the past when having the experience of remembering, or shortly before. On the other hand, making stored information available for retrieval is a process that starts as soon as we have an experience, and is ongoing until the experience of remembering and afterwards. This is a process over days, months or years which determines what gets remembered; the constructivist is concerned with a process that takes minutes which determines how the experience of remembering is generated and experienced.

What is retrieval? It is a process in which a system supplies an item or set of items that relate to a query. Queries can be explicit—such as when I deliberately search my memories to find my brother's girlfriend's name—or implicit—such as when seeing a person's face cues a memory of their name. Retrieval critically involves two processes working together. In the long run, stored information must be structured so that important things can be retrieved. In the short run, search has to operate on stored representations successfully to locate relevant stored content. These two processes are extremely hard to separate empirically, and so I will not attempt to do so here. Instead, by retrieval, I will refer to the joint operation of the two.

Retrieval is a functional bottleneck for memory, a step in the process to which many failures and successes can be traced. This holds for both animals and artificial systems. In using this fact to argue

for the normative centrality of retrieval, I will assume that one of the theoretical virtues that makes for a good account of the function of a biological system is that it can explain common successes and failures.¹

Research on human memory suggests that retrieval is a major source of memory failures (Sweatt, 2009). One way of seeing this intuitively is to think of all the times you could not remember some fact that later came to you easily in a different context. In computer memory systems, where it is fairly easy to tack on another server or hard disk or some other storage device, the retrieval problem is even more prominent; consider how much improvements in web search facilitate our use of information. One could argue this is just a feature of the peculiar strengths and weaknesses in our technology, but it is significant insofar as it aligns with the challenges found in biological memory systems. The fact that retrieval shows up as the factor which separates successful from unsuccessful memory systems across the board suggests that retrieval is a challenging computation even at a high level of abstraction.

Strategies to enhance retrieval are central to memory expertise in humans. The Method of Loci (MoL) is a strategy used by memory experts as far back as Roman times (Cicero, May & Wisse, 2001), and notably described by A.R. Luria as used by the brilliant mnemonist Shereshevsky (Luria, 1968). An example of this technique would be memorizing a list of unrelated names by visualizing placing each name in a familiar location on a walk around my childhood neighborhood. Then, to retrieve the items, I would again visualize my walk through the neighborhood, this time picking up each name one by one. The MoL is not just used by experts but can be taught to older adults to improve memory performance (Dagleish et al., 2013; Gross et al., 2014). The MoL is a retrieval strategy because it involves structuring one's memories so that they can be effectively retrieved later, and prescribes a specific strategy for that retrieval process. The symmetry between how the memory is stored (in a visualization of the childhood home) and then consequently retrieved is striking, and illustrates the property of retrieval I discussed above: effective retrieval involves tight cooperation between how information is stored, and how it is searched.

Another line of evidence comes from retrieval pathologies such as enhanced retrieval in Post-Traumatic Stress Disorder (PTSD). While many extant philosophical discussions of memory (Bernecker, 2010; Michaelian, 2011b) refer to the case of H.M. who had traumatic damage to his memory system which affected storage, a far more common memory symptom, typically experienced in PTSD, is intrusive memory. In these memories, a traumatic incident is remembered vividly and often. This is not because the traumatic memory is often relevant, but because something is wrong with the retrieval function which privileges a certain set of memories (the traumatic ones) without good reason. That is, these memories interfere with thriving and so are defective in a biological sense, but they are also defective in another sense; through retrieval of a memory, they present some event as relevant even when it is not.²

So if retrieval keeps showing up as the key step in memory, we should expect a good theory of memory function even at an abstract philosophical level to either explain why this is or provide a basis for such an explanation. An objection to this way of thinking about memory function might be that for ideal epistemic agents, retrieval is inessential. Perhaps these agents can access all their stored representations instantly in parallel, or maybe the idea of access as a process itself presumes the system in question is non-ideal.

¹ I present a related argument in a somewhat different context in Aronowitz (2018).

² Another disorder that may involve defective retrieval is obsessive-compulsive disorder (OCD). Patients with OCD at least sometimes present with memories that are incredibly detailed but with details that are somewhat disconnected from a reasonable sense of what is important to recall. Of course, a more detailed discussion would be necessary to establish the plausibility of this way of looking at memory in OCD.

I concede that under some notion of the ideal agent, this objection is correct. God probably would not need a function to get information from his memory system. One response is to be more specific about what I mean by an ideal agent: a creature that carries out the same operations as we do, but is not limited absolutely by computational difficulty. That is, this kind of agent is capable of doing any algorithmically specified computation, but not, say, processing uncomputable functions. I am also presuming that we can still talk about efficiency for such an agent. This might seem odd, but it strikes me as an assumption shared by classic Bayesian epistemology, as well as other algorithmic but idealized frameworks. For instance, the Bayesian advocates for conditionalization, but does not take herself to be obligated to rule out the procedure of first transforming the credences linearly, then conditionalizing, then re-transforming. Presumably this is because that would be no better, and much less efficient, than conditionalizing.

In my case, I am appealing to efficiency to support the necessity of retrieval for the ideal agent in the following way: were the agent not to intelligently store and structure her representations in order to optimize them for retrieval, she might still be able to function at the same level by employing an extremely complicated and computationally costly retrieval mechanism. However, efficiency would best be served by having a retrieval function and representational structure that work together. One further thing to establish is the criterion for successful retrieval. I will assume that this consists in two features: (a) truth and/or accuracy, and (b) relevance and completeness.³ This means that the information accessed in response to a query must be on topic, not omit anything of clear importance, and of course, be true or accurate. One thing you might expect to see in addition to (a) and (b) is something like faithfulness, or accuracy to the past representation, rather than to a state of affairs in the world. However, faithfulness is not needed to characterize retrieval, and in fact, as I will discuss later, is often at odds with retrieval.

So I have argued that retrieval is integral to successful epistemic functioning in memory, and that successful retrieval consists in storage and search that makes available true and relevant information.

3 | SLEEP AND MEMORY

Animals, including humans, seem to be fairly successful at solving the retrieval problem. We are exposed to an incredible amount of information every day, and yet manage to pull out important facts to highlight and be reminded of later. This suggests that looking at animal memory system might tell us something about how the problem can be solved, and what systems that solve the retrieval problem have in common.

In this section, I will aim to ground two claims: (a) functionally, information stored in animal memory systems tends to acquire “model-like” features over time, and (b) these features are developed, updated and maintained primarily or at least to a significant degree by the memory system. I will examine what it means to be model-like in Section 4, but for now, the idea is to let the data guide us toward a sense of what these information changes have in common. (b) is necessary to rule out the reasonable hypothesis that on-line deliberation is doing all the interesting computational work and memory is merely its storage receptacle.

³ In computer science, relevance/completeness is referred to as the recall/precision trade-off; these trade-offs both refer to the fact that the more inclusive search results are, the less they can precisely pinpoint connections between the query and the retrieved items.

3.1 | Predictions

I will now turn to the connection between sleep and memory in animals. I will focus on sleep because what is at issue between the various theories is whether the memory system itself performs information transformations. Because we neither deliberate nor perceive during sleep, informational changes that happen in memory during sleep can be attributed more cleanly to the memory system itself as opposed to deliberation or perceptual processing. So sleep takes some possible confounding variables out of the equation.

Every theory on the continuum from preservationism to constructivism can accommodate sleep having some effect on memory. What happens in memory during sleep, however, could be understood under very different hypotheses depending on the background theory:

Tagging hypothesis: Deliberative systems tag information changes which are carried out by the memory system (but neither designed or planned by the memory system).

Index hypothesis: Memory systems build an index during sleep (to a lesser degree during waking) which is used to pull data from a long term, relatively inert, storage system.

Model hypothesis: Memory systems build and alter models during sleep (to a lesser degree during waking) which are used directly in retrieval.

Constructive hypothesis: Memory is an on-line process of imagination and so should not be implicated in sleep except in the way that all cognitive operations tend to degrade in conditions of sleep deprivation.

The tagging hypothesis about memory is a way of articulating that memory is informationally inert and therefore fits with preservationism. On the model and index hypotheses, memory performs informational operations, analogous to the kind of curation introduced in the museum analogy. I will not get into the differences between these two views until the next section; instead, I will focus on what they have in common. Namely, both predict that animal memory systems, to be successful at solving the retrieval problem, should alter stored representations to make them simpler, more cohesive and structured according to patterns and generalizations. I will argue that the data strongly favors both the index and model hypotheses over the tagging and constructive hypotheses. Note that while constructivism and preservationism are two ends of the spectrum, they generate virtually the same prediction here, since both minimize off-line information processing in memory.

3.2 | Data

While there has been a lot of behavioral and higher-level work on long-term changes to memories (most famously by Loftus (1996)) and other phenomena which might support my view, in this section I will concentrate on low-level neuroscientific data.⁴ Additionally, the model I am presenting is meant to be a reinterpretation of a swath of research that investigates the influence of environment and more generally prior beliefs on memory.⁵ These models often either give a fully general,

⁴ This is in part because these kinds of structural question can be answered most directly by looking at mechanisms, but also because the findings I discuss in this section have gotten less attention in the context of these abstract debates about memory function.

⁵ It is worth noting that some models—such as Zhang and Luck (2009)—explain data of the kind I'll discuss without any direct influence of priors on memory (in their case, by a combination of straight remembering and on-line guessing). However, this is plausibly due to differences between working and long-term memory.

domain-neutral theory (Anderson & Milson, 1989; Hemmer & Steyvers, 2009) or focus on small-scale, particular schemas (Tse et al., 2007). My project is to give a re-description under which these results are picking out pieces of the same underlying phenomenon.

At a minimum, it is clear that in humans, rodents and drosophila, sleep is critical for memory consolidation. The task of this section is to argue that what happens during sleep is likely to be a process of organization and prioritization. If animal memory is a model system, then these are precisely the features we should expect. By contrast, other views do not predict this. That suggests an explanatory advantage in favor of the model view.

Several patterns have been well established in the research on the link between human and rodent memory operations and sleep, including the following: enhancement for weak rather than strong memories (Diekelmann, Wilhelm & Born, 2009), preferential enhancement of goal-related memories (Walker & Stickgold, 2004), and enhancement of temporal ordering information (Drosopoulos, Windau, Wagner & Born, 2007; Griessenberger et al., 2012). All of these facets indicate that informational transformations are occurring. Another line of evidence against the tagging view is the two-way connection between memory enhancement and sleep—in studies in drosophila, researchers found that increasing the memory demand in the environment increased time spent asleep, directly stimulating brain regions responsible for memory increased sleep (Dissel, Melnattur & Shaw, 2015), and enhancing sleep chemically reduced memory deficits (Donlea, Thimgan, Suzuki, Gottschalk & Shaw, 2011). These interventional studies indicate that at least in flies, there is a genuine causal connection between memory and sleep.

Let's consider one finding in some more depth. Drosopoulos et al. (2007) looked at consolidation of temporally ordered information in humans in a sleep and control condition. They used repeat sequences to separate temporal ordering effects from more general ordering effects. Consistent with other work (e.g., Diekelmann et al., 2009) they found sleep only strengthened the forward temporal sequence and not the backwards one—that is, subjects who had slept between training and testing were no better than those who had not at the task of naming which cue came before some cue that was shown to them. However, they did significantly better than the non-sleeping group at naming which cue came after a cue that was presented to them. A second finding was that weak associations were strengthened more than stronger ones—if an associated was less reliable during the training session, it was more likely to be improved in the testing session.

This and other behavioral effects of sleep have increasingly clear neural correlates, at least in rodent models. For instance, Bendor and Wilson (2012) taught rats an association between a tone (left or right), and the location of food (at the left or right end of a track respectively). Following training, the authors recorded from hippocampal neurons. Critically, these neurons represent a spatial map of the environment, and go through “replay” during slow wave sleep, in which patterns of activation experienced during waking are repeated multiple times in the same order during sleep. They played three sounds (the left and right tones used in the training phase, and an unconditioned control tone) during the two sleep phases as well as waking. Playing the left tone during Rapid Eye Movement (REM) sleep (but not slow wave sleep or waking) enhanced hippocampal replay in the left region of the place field, that is, the part of the neuronal ensemble that represents the left half of the environment. The same was found for the right tone. Further, playing either tone lead to behavioral changes; the rats were more likely to make errors in favor of the side on which the tone was played during REM sleep. We have copious evidence of a correlation between hippocampal consolidation during sleep and changes in waking behavior, but this experiment suggests a clear causal connection; by biasing signals during REM sleep, the experimenters biased the neural patterns at consolidation, as well as the behaviors during subsequent testing. That is, not only does hippocampal consolidation

parallel the information transformations I have been talking about, but it is likely that changes in consolidation drive changes in behaviors.

These two lines of evidence (informational effects and causal connection) count against the tagging view in the following way. Since the tagging view says that activity during sleep is merely carrying out the instructions generated during waking, it should predict that stronger associations have an equal or greater sleep boost than weak ones. The causal studies indicate that there is a two-way relationship at play, since enhancing sleep enhances memory and learning.

The model view predicts a computationally taxing and critical period of consolidation, since models have to be assembled and designed in a non-trivial way. One line of evidence in support of this claim is sleep-linked memory triage. This is a broad category that encompasses many ways in which information acquired during the day is altered and consolidated during sleep. In selective item consolidation, subjects tend to forget items which do not fit into a pattern (for fMRI corroboration on this point see Rauchs et al., 2011), and reinforce those that do. In multi-item integration, a new stimulus that fits into an existing pattern will be integrated with that pattern, and in gist extraction, a pattern will be generalized leading to confabulating items which fit in with the pattern but were not presented. To simplify a little, it seems that consolidation during sleep is important for fitting new memories into existing patterns, and extracting new patterns.

The phenomena of information transformations like gist extraction are compatible with a wide range of memory structures, though the idea of active forgetting (e.g., in Rauchs et al., 2011) makes some trouble for a strict preservationist picture. What I take to support the model view is that this consolidation takes place during sleep. If reasoning or other on-line processes were responsible for gist extraction and generalization, we would expect items that did not fit into a pattern to be recalled less even without a period of sleep before testing. However, there is a significant sleep-dependent boost for each of the three effects I have discussed (Walker & Stickgold, 2004). One explanation for this draws on the neurochemical environment during REM and transition-to-REM (TR) sleep. Poe et al. note that the low acetylcholine and high nor-epinephrine during these phases facilitates long-term depotentiation, the reversal or dissolution of long-term potentiation, which is often described as the neural correlate of associative strengthening (Poe, Walsh & Bjorness, 2010).

Sleep-linked processing is by definition off-line, and the neural correlates of these consolidation processes indicate areas associated with memory, rather than other reasoning or learning processes (Poe et al., 2010). So it seems that these operations are operations of the memory system itself. First, this evidence is predicted by the model view, since consolidation will involve integrating new evidence into existing representations, which has the automatic consequence of generating inferences along the lines of the confabulation cases, where the subject claims to have seen an unseen stimulus "EFG," which follows the three-letter pattern of the stimuli which they actually saw.

Stepping back from the details of memory consolidation, another line of evidence appeals to the division of the memory system into episodic and semantic components. Semantic and episodic remembering are two ways of representing which can share some representational content (for instance, I can semantically remember that Fatima was the daughter of Mohammad, or I could episodically remember my teacher telling me that Fatima was his daughter). Whether I prefer one form over the other depends on which features of the information I deem relevant. If I doubt that my teacher is a reliable source, I might want to remember that it was him who told me that fact just in case I get more evidence about his reliability. In other words, the choice of format, just as with models more generally, is relevant for determining how that information will change over time. The

challenge for the model view is showing that these large scale patterns are replicated at a lower level.⁶

What does all this data tell us? It paints a picture on which after sleep, our memories are more organized, combined and structured according to patterns, and sometimes expanded and elaborated in accordance with these patterns. This change is not just observed behaviorally, but we can follow the neural correlates of the representations as they change, and make interventions to predictably affect the outcome of this process. These changes are at odds with the tagging and constructive hypotheses, and thus in tension with preservation and constructivism.

4 | TWO RETRIEVAL SYSTEMS

What does it mean to structure information according to patterns? In this section, I will discuss two conceptual views of memory that capture the kind of information changes observed during sleep: model systems and index systems. For each, I will define it, relate it to the retrieval norm, and then give some examples of that system in the world.

4.1 | The model system

A model system is any information processing system that builds, evaluates and otherwise processes models. This is not very helpful—the real question is what is a model?

Informally, I will take models to be representations that work via mapping dimensions of the system being represented (which I will call the object) onto dimensions of the representational domain (which I will call the medium). These dimensions cannot just be points; what differentiates a model from, say, a group of sentences or a table of facts is that it encodes not just the current state of the object, but also regularities, laws or expectations about how features of the object might change (i.e., dynamics). So for instance, a plastic model of an atom counts as a model because the way that the plastic beads move around on the rings encodes dynamics of the movement of electrons across the levels. Now, note that some features and dimensions of the plastic atom model do not encode anything about atoms; the fact that the rings are painted in colors from blue to red, for instance, says nothing about any property of atoms. Conversely, many features of actual atoms are not encoded in the model, such as their spatio-temporal locations. So a second property of models is that they are simplified: they function to encode only some of the total features of the object.

Further, the model atom represents the dynamics of the electrons via the relationship between its plastic parts. The final property which I will ascribe to models is that they are relational. If we had a model of the atom and a plastic model of a zoo sitting next to one another, their union does not form a new model. And likewise, within the model atom, were we to separate the various parts based on which particle they represent, this breaking down of the model would result in information loss. This is what makes the atom model a model, but the union of atom and zoo models not a model.

One caveat: I am not aiming to give necessary and/or sufficient conditions for the common-sense category of models. Instead, I am just presenting criteria for being a certain kind of model that will then explain why models are suited to the retrieval task. As a class, models are representations which have these properties:

⁶ Buzsáki and Moser (2013) suggest that the semantic/episodic divide might be grounded evolutionarily in the allocentric/egocentric distinction in spatial navigation. If this is right, since spatial representations are quite obviously model-like, there is a good reason to think the memory system which evolved from them shares this property. See Colgin, Moser, and Moser (2008) for evidence that spatial representations are relational.

Domain-specific: the representation employs structures or uses a medium which is suitable specifically for the kind of thing it is used to represent, as opposed to being generically suitable for representing anything.

Relational: all ways of dividing the representation into an unordered set of local representations will result in a loss of information. This is because some information adheres in the relations between parts of the representation, and so is lost when we separate it into parts.

Simplified: the representation has fewer dimensions than there are in the information representable about the set of objects it depicts. (Here dimensionality is understood in the information-theoretic sense to mean a feature space: for each feature or aspect of the data, we add one dimension to the model.)

What is the difference between a model and a network? In the case of a relational database, for instance, we talk about networks having some of the features I have ascribed to models. I take the difference to be that there is a model-like *manner of representing*, whereas networks are ways of *implementing representations*.

Chalmers (2004) describes manners of representing as the impure content by which the pure propositional content that the cat is on the table can be represented in a visual manner, or a conceptual manner and so on. These manners occur at a variety of levels of description, such that one representation will usually be characterized by several manners of representation. This does not entail that manners of representation are informationally neutral: that is, they sometimes constrain the content in a context. Imagine we have an auditory and visual representation which have as their content the same set of propositions about the cat on the table. Now, if we add the proposition that the table is on top of the rug, the two representations may update in divergent ways, adding different entailments to their content. For instance, in the visual representation, it will surely follow that the cat is above the rug. This might not be part of the content of the auditory representation. So while different manners of representing can have the same content, they will not in general have the same dynamics, or changes in content over time and the addition of new information. Ways of implementing representations, in contrast, are genuinely informationally neutral.⁷

So the claim that there is a model-like way or manner of representing is tied to the fact that models encode dynamic inference patterns. Even when the same static content is represented in two different models, the total content in each case will tend to change in divergent ways when updating on the same evidence.

Putting all this together, what it means to say that memory systems are model systems is that they build, manipulate and maintain simplified representations with a domain-specific functional architecture. This is an operation over multiple models, some of which may overlap. How many of these representation structures there are and how they fit together will depend on the agent or system in question, but the common thread which I will now argue for depends on the aforementioned characteristics of models along with the idea of using multiple representations in parallel.

Now that we have a working theory of a model system, we can ask: why single out this system as a solution to the retrieval problem? Here is the idea: in order to be effective under the beginning of

⁷ Of course, just like any feature of the implementation of some content, it can constrain the content; the machine might not be big enough for some degree of information, for instance. But in general, two identical representations can be stored in totally different ways without this effecting their dynamics.

inquiry conditions (a)–(c), retrieval systems need to be able to deal with an influx of new information that bears on previously learned information. To do so, they need to be dynamic—capable of using this new information to update and change the current representations. Second, the system needs to be able to extend current information to new cases—some queries will require extrapolation or interpolation, rather than just reading off what is already recorded. I will call this ability to extend to new cases prediction, though it need not be about the future. Third, the representations in the system need to be searchable. These are meant to be central attributes of successful retrieval systems, but not exhaustive.

The relational features of model systems allow for dynamics, since isomorphic ranges essentially encode a series of behaviors, and given that the medium itself is bound by rules (say, spatial rules), the rules on the medium will generate rules for the representation. For instance, take a model of an argument in a conceptual map on a spatial medium. The isomorphism here is between conceptual nearness and spatial contiguity. So it follows that the conceptual nearness, according to our model, will always be symmetric. Intending to represent the concept of “art” as closer to the concept of “dreams” will result in the symmetric revision that “dreams” will be closer to “art,” and most likely many more changes.

The simplified nature of models enhances prediction. This is because simplicity is known to increase predictive power by pushing the model away from overfitting. Here are Hitchcock and Sober (2004) on this point: “For the complex model to have the higher estimated predictive accuracy, it isn’t enough that it merely fit the data better than its simpler competitors (that’s pretty much inevitable); it must fit the data sufficiently better to compensate for the loss in simplicity it represents.” And it is easy to see why simplicity enhances search.

Finally, the domain-specific nature of models supports each of the characteristics. Domain-specific representations can encode more regularities, since many regularities hold only within domains and not in general.

The reason to have multiple overlapping models can be derived from the second part of the environmental condition—that the agent expects to get a lot more information in the future. With more and more information, the models the agent currently has might become quickly inaccurate. One reason to have multiple models is to be risk-averse. Part of the distinctive role of memory is to provide a database of facts that can be re-purposed even if the agent’s fundamental assumptions and beliefs have changed. Multiple overlapping representations help fulfill this role by allowing the agent to have a range of representations with different dynamics. Episodic memory, for instance, seems that it is slower to change than semantic memory; this fact may account for why folk psychology seems not to treat episodic memories as completely static. So this allows the agent to use episodic memory as a source of recovery if their semantic memory needs to be radically revised—this means the model system retains prediction, dynamics and effective search in cases of recovery or goal drift.

Overlapping representations allow the agent to use multiple functions; for instance, if I’m having trouble remembering in general what kinds of things people give as housewarming gifts (a semantic task), I can switch to episodic retrieval and think up examples of things people have given me as housewarming gifts. This is particularly true if the search functions have exponential cost and are reasonably independent; in that case, adding a new function will massively increase the likelihood of retrieval. In addition, different manners of representation allow for sequential versus random access retrieval, so the agent can use a flexible retrieval strategy which exploits the advantages of both methods.

4.2 | The index system

In the model system, the entirety of stored information is transformed and used in models. The index system, in contrast, keeps the underlying information in unstructured form, while using a dynamic index to alter what is retrieved over time. As I will discuss in Section 4.3, these two systems exist on a continuum. While I have not found anyone directly defending the index system as a model of memory in the philosophical literature, I think it captures an intuitive idea: why can't we just work with unstructured stored data by means of a very intelligent search system? I will build this thought into a competitor for the model system in this section.

An index system has two representational components: a (mostly) unstructured dataset, and an index. Processing in this system involves constantly updating the index, and retrieving items from the dataset via the index. Essentially, the index functions as an intermediary between the query and the dataset. The index itself can be relatively simple, or very complicated, consisting in multiple ways of retrieving the data and highly structured clusters or properties.

While an index in the back of a book is usually an alphabetical list of terms with a link to a page number in the text, indexes for more complex retrieval systems tend to be inverted. This means that instead of a list of terms, we have a list of documents. Each document has a set of entries associated with its content. Unlike a table of contents or index in the back of a book, an inverted index represents the documents through their contents rather than representing some content through its location in documents. A simple method for indexing is to represent documents as vectors, where each cell records the place of the term in the string. For instance, the sentence "I should store this now" can be represented like this:

$$\begin{bmatrix} \text{I} & \text{D} & \text{s} & \text{h} & \text{o} & \text{u} & \text{l} & \text{d} & \text{s} & \text{t} & \text{o} & \text{r} & \text{e} & \text{t} & \text{h} & \text{i} & \text{s} & \text{n} & \text{o} & \text{w} & \text{I} \\ 1 & 2 & 0 & 3 & 4 & 5 & 1 \end{bmatrix}$$

In this case, I recorded all of the data in the string, but often, I only want to index some of the data: for instance, I could index all of the titles of articles in a news archive.⁸ Typically, because even an index this light can become expensive, we would try to reduce the dimensionality in various lossy or lossless ways depending on the application.⁹ Then, I could use the simple representation above to look for articles that use the phrase "should store" by looking for numerals n , $n + 1$ in the second and fourth columns. For every time that pattern appeared, I would use the entry in the first column to bring up the document containing that term in the headline.

This kind of system is widely used in web search. I will focus on index systems for web search over natural language data; similar systems are used for searching and indexing images, but natural language data is complicated enough to see the features of indexing that are important for my purposes.

Essentially, all web search operates via indexing; users put data on the web in formats that range from semi-structured (such as a Wikipedia entry) to unstructured (such as an image), and search engines crawl through that content and index it. Cutting edge indexing procedures for engines like Google involve combinations of hand-coding, intelligent use of user input, machine learning and so on. User input can take the form of explicit feedback but more often involves things like taking the fact that they went several pages into the results before clicking as a failure (see Liu, 2009 for recent work on making the ranking process more intelligent).

⁸ Data structures such as tries (Fredkin, 1960) that store the entire dataset are sometime classified along with indexes, but for my purposes an index must be an intermediary between the query and a further dataset.

⁹ For instance, using singular vector decomposition (SVD).

The index part of the index system shares some of the features of a model system, even though the underlying unstructured data is different. The index is usually simplified—unlike the index used above, all of the information in a document is almost never represented. Indices are often relational—an easy way to compress an index is to omit strings shared by all documents, which has the result that the representation of any particular document is determined in part by what other documents are being stored. However, most indices are not domain-specific. For instance, an index used for web search handles all kinds of natural language information. In part, this must reflect the practical use we put indices to. However, a genuine difference with models is that an index can be highly content neutral, such as the example above, or very content-specific.

4.3 | Comparing the index and model systems

How does the index system solve the retrieval problem? It updates the index frequently and intelligently, in a way that involves not only adding entries, but changing up the structure. This allows for effective search by keeping the index sufficiently simple and making sure it represents or draws on relevant features of the data. For example, the index to a database of news articles should probably encode (among other things) the publication dates of those articles. But even better would be an index that had a hierarchical structure. For instance, it could embed an index that separated articles by sentiment under one that characterized them as movie reviews or sports articles. As it turns out, many indexes are organized this way—not unlike tags and sub-tags on a calendar app.

The step I made in the last couple sentences of the preceding paragraph is a blueprint for the argument of this section. As we ask indexes to do more and more difficult tasks, and grant that they have the necessary computational capacities, an index system as a whole starts to function more and more like a model system.

What makes web search different from human memory? One difference is that while the data in web search is added by users who all have different goals, in our memory systems, the same agent adds data and categorizes it. Even if you take a very modular approach to the mind, there is still both a set of interests that characterize the person as a whole, and a good deal of calibration between cognitive faculties—even if this calibration is unsupervised and unconscious, or even built-in. In web search, content is often added in an adversarial way. For instance, most companies who have web pages want them to appear as high on the ranking for any search query as possible, whereas the search engine wants only results that are relevant to users (or at least that users will click on).

This difference explains why web search must use indexes: the task of structuring the data itself, rather than superimposing an index with structure, is much harder given this array of goals than for a human brain with a cooperative and calibrated input system. But even in current search engines, you can see the intermediate stages between an index system and a model system. Think about how many times you put in a search query and found your answer just by looking at the results page without clicking on any of the retrieved items. This is a case where the function of retrieval is being satisfied just by the index representation, instead of the index acting as an intermediary. How can the index itself fulfill this function? By being representational, relational, simple and so on.

An interesting case here is the use of higher-order features of the data, rather than the data itself, to answer queries. Xia, Cheng, Prabhakar, Lei and Shah (2008) present a way of indexing on which the mean and variance of the data is used instead of the values in the data itself. Rather than being computed on demand from the raw data, the mean and variance are updated consistently in the index and subsequently used to answer queries. This model is intended for situations with lots of fluctuations in the data around a fairly stable mean; the computational challenges of indexing on-line around this kind of data are significant, so using this higher-order feature instead decreases costs and more

significantly, increases *reliability*. For instance, location-based services like smartphone navigation involve a constant and variable influx of data—Xia et al.'s system uses the mean and variance of this information rather than tracking more detailed information from the raw data, and by so doing can provide better and more stable guidance. Essentially, this index is working like a model, and so it outperforms more traditional indexes.

Thus there is reason to believe that as we relax the computational limitation on an index system like web search, and allow it the power to incorporate more of the content into the index in a hierarchical format, queries and other user demands will be satisfied more and more by the index itself rather than using the index as a window into the raw data. Modeling can become more and more useful even with no change to the goal structures, as the Xia et al. case shows—there we just adjusted the relative complexity of the various computational stages.

Of course, I am not suggesting that it is optimal to discard the data. I would agree with most epistemologists that one should never give up information if it doesn't cost you anything. But the data will become less and less relevant to the day-to-day operations of the system. Thus, functionally speaking, as we approach negligible computational limitations, the index system will function just like the model system. If you're not particularly interested in the ideal agent, it is also worth noting that the relevant limitations in the connection between data and its analysis are not particularly applicable in the case of memory in a biological organism.

On a more empirical note, the mammalian memory systems described in Section 3 resemble a model system more closely than an index system. One relevant finding is confabulation. The increase in confabulation following sleep, such as the memory that I saw “EFG” when in fact I saw a set of other alphabetical strings in the same pattern, is not accommodated well by the index view. As I have noted, generalization and gist extract during sleep leads to the creation of new, confabulated memory entries. Confabulated memory entries are at least sometimes produced by the memory system. If memory is an index system, the index is not responsible for generating content because it does not represent content on its own but merely acts as an intermediary. So on the index view, sleep-linked confabulation must either be explained away, or a new mechanism posited to explain the addition of content to the index and the storage system, respectively. By contrast, on the model view, the confabulation results are expected as a consequence of simplification. For instance, a natural simplification of the sets of letters is something like “I saw a bunch of strings of three consecutive letters”—and of course, the confabulated string fits just as well under this heading.

5 | EPISTEMOLOGICAL CONTEXT

Here I will survey some views on memory in the epistemology literature that contrast with the model view. I will assume the framing of the retrieval problem, and the environmental conditions (a)–(c), and argue that these other options fail under those assumptions to give a satisfactory account of the contribution of memory to knowledge.

5.1 | Causal theories

The first class of extant views are causal and preservationist theories. While they have key differences, all these views subscribe to the following:

Origination principle: For every memory with content p , there is a unique original mental act with content p^* to which that memory can be linked (which forms the grounds for identifying its justification and/or content).

This will be the target of my arguments, but first I will discuss the views themselves.

For Bernecker, who starts with a conceptual analysis of the individual states of memory, memories are mental representations which share a continuous internal causal connection with a sequence of states of the same content, tracing back to a single incident of perception or some other kind of doxastic act. In the latter case, for instance, my memory of the concluding step of a theorem will relate back to my realization of that step. In general, the original incident will be a case of coming to know, coming to believe, or coming to acquire the information which I will call the content of that incident. When the original incident is a propositional attitude, the content is simply the content of that attitude; if the incident is non-propositional, which will be the case under some views of perception, the content is something like the proposition that corresponds to the non-propositional content. For the sake of simplicity, I will focus on the case where the original incident is a propositional attitude.

It is this original incident that gives the memory its content: “memory contents are fixed by the past environments and remain unchanged until some later moment of recollection” (2010, p. 169). On Bernecker's view, when I am slow-switched from Earth to twin-Earth, my memories with the content “water” will refer to water if the original incident was pre-switch, and twin water if it was after the switch. This view falls under a more general class of causal views, where what makes some state a memory depends on a causal connection between that state and the other states which make up the memory trace. Causal theorists are not all committed to sameness of content among the states in the trace; Bernecker allows for entailments of the original content to be added to the later content, for instance, and Michaelian (2011a, 2011b), in earlier work, before rejecting this view to turn to constructivism, allows for change and reconstruction so long as there is a “sufficient” degree of similarity between all the contents in the trace. But in either case, these views all clearly presuppose the origination principle.

Causal views of memory have a few advantages. I take it that they are quite intuitive, and express something like our common-sense notion that memories are preserved contents from the past. Causal views also allow for a neat and simple treatment of memory content; since each memory can be traced back to an initiating epistemic act (for instance, my memory of seeing an apple traces to the act of me seeing the apple), it simply inherits the content of that act. Thus, memories refer to the past and restrictions on perception, imagination or other doxastic states which might be initiating events carry forward to restrictions on memory. If I am not in the position to form beliefs about the paradoxical circle-square, for instance, I will not be able to remember anything about it either.¹⁰

These causal views are a natural fit with a position about the function of memory: preservationism. As presented in Burge (1997), an agent is entitled (or *prima facie* justified) in holding a

¹⁰ Campbell (2002) offers an alternative to the causal view. He argues that since many memories can not be traced back to an initiating event, it is more plausible that memories are about possible past instances than actual ones. For instance, when I remember my grandfather's gas station, it seems implausible that this entails that I am remembering a single instance of being at the gas station. Instead, I could have taken elements from many visits and compiled them into a single representation. Campbell proposes that this representation be treated as a representation of a possible way the gas station could have looked to me. This account avoids the objection to the causal theory that our memories rarely seem to meet causal criteria, but it is vulnerable to an objection about the accuracy of memories. What does it mean to truthfully depict a possible past incident? I will set Campbell's view aside for now since I can imagine extensions of it that would be consistent with my conclusions, and others that would not be.

remembered belief if the belief represented in the memory was originally justified. More generally, if the original incident was of epistemic value (if it was a case of knowledge, justification, responsible-ness, etc), memory functions to preserve that status. One could extend this view to infer a norm for the memory system, in the following way. First, epistemically well-functioning memory preserves justification. Since justification is inherited from the original belief formation, memory ought to hold on to beliefs in as close to the original condition as possible. It is up to the belief formation and revision system, in other words, to come up with justified beliefs, and the memory system has the more passive role of preserving that justification.

In response to objections that preservationism is a bad fit with psychological evidence, Michaelian proposes a weakening of the thesis on which “for any finite cognizer, a certain pattern of forgetting is necessary if her memory is to perform its function well... by eliminating ‘clutter’ from her memory store, this pattern of forgetting improves the overall shape of the subject’s total doxastic state” (Michaelian, 2011b, p. 399). On his picture, the ideal agent would preserve all of their memories, but for agents with limited capacities for storage, the norm is to store up to the point when storage inhibits retrieval: “Even where the subject actively updates her beliefs, the record that underwrote an outdated belief can continue to lower the reliability of retrieval” (p. 412).

In summary, the views outlined above are in agreement about several points. First, they define the content of memory or the norms governing memory by reference to a single original incident which is not itself a memory. This requires the origination principle. For Burge, this original incident is the original belief (or perceptual belief etc.) which is either justified or not. For causal theorists like Bernecker and Michaelian, it is the incident which fixes the content of the belief. It is natural, but not strictly logically necessary, for these incidents to be the same. A second similarity is that memory is conservative in nature, functioning to maintain or sometimes repair content or justification; while some of these views allow for changes in content, these changes are restricted by the original incident and there is no positive norm provided for memory change. Note that this conservatism depends on the first: without a fixed idea of the original thing, there is nothing to conserve.

My first objection to the origination principle is just that it is unmotivated—what is so special about the original content? A reply might be that we have direct contact with the original contents, or they have a greater degree of justification. There might need to be a system to maintain inferences off of these core contents, but that would not be the job of the memory system.

Note that because original incidents can be perceptions, realizations, emotions, and all kinds of other mental acts, there will need to be a privileged content in every case—and to avoid a heterogeneous and disjointed system, these contents will have to either be the same or easily translatable between each category. If we look closely at what makes up a remembered episode, I doubt that there will be a natural unit which forms the secure content; instead, while some of the information will be more or less noisy, there will not be a difference in kind. And in fact, sometimes we might use many memories in the origination-principle sense to generate a extrapolated content of which we are more certain than any of the memories. I have argued throughout this paper that we should expect this to be part of ideal memory functioning—for the system itself to generate new content via abstraction, simplification and so on—and yet this content would not count as remembered given the origination principle. Here, Michaelian would object that causal views can still account for some content change and addition. But this is only in non-ideal agents.

Overall, the problem with the origination principle is that it imposes a means of individuating content on the memory system. In contrast, the considerations I have offered support an epistemic

advantage if the memory system itself is responsible for building and changing representations. Memory faces different challenges than other systems, and has unique data needs. Even if these new structures can be described in terms of inherited content, as Michaelian wishes to do, doing so adds nothing to the epistemological theory. Insofar as epistemology and theories of content should go together, origination theories cannot work.

5.2 | Memory-as-testimony

This section is concerned with a view that does not rely at all on the origination principle. In contrast, it treats the epistemic value of content supplied by memory as equivalent to testimony. Given the wide variety of views on testimony, this theory of memory could work in a wide variety of ways. For simplicity, I will assume the following view of testimony, which Barnett (2015) calls the naive view: “when a source of testimony tells you that p , what you learn first is not p itself but instead merely the fact that the source says that p ”. Barnett goes on to qualify that this is an internalist view in that it depends on your justification in believing that your source is reliable, rather than the source actually being reliable, but I will stay neutral on this point and allow either version to count as the memory-as-testimony view.

One relevant issue for the testimony view is that the retrieval problem that I have argued is central to memory is outside of the normative scope of the view. Why? Because, just like in the case of ordinary testimony, when we are asking the up-taker of the testimony to justify her reliance on it, this view is about how and when the rememberer is justified in relying on the information served up by her memory system. The retrieval problem is about how to store and modify information in order to offer the correct entries up in response to a query; that is, in the analogy, good retrieval is like being a good testifier, not a good up-taker of testimony.

Of course, it is coherent to ask what a good testifier looks like. I would like to note here that answering this question is not part of the analysis offered by extant memory-as-testimony accounts. More problematically, it is the focus of almost none of the testimony literature itself. And in fact, this may be because epistemology should be relatively permissive about what constitutes a good testifier. It does not seem irrational for me not to fully optimize my communication to serve your epistemic goals, let alone to prepare my storage systems for years in advance in order to offer up information which will be both accurate and also relevant to your needs. Even if this would be excellent social behavior, it seems excessive to suggest that it would be rationally required. And yet, this is exactly what we would have to do to maintain both the testimony analogy and the importance of the retrieval problem. In conclusion, I have not argued that there is no useful analogy between memory and testimony, but merely that the epistemic analogy between memory and testimony seems unable to explain anything about how to solve the retrieval problem.

6 | WRONG LEVEL OF ABSTRACTION?

You might think everything I have been saying is sort of reasonable for structuring a machine, but not of normative importance. How could something as low-level as manners of representation within the memory system bear on epistemic rationality?

In response, I emphasize that it is not the details of the model system doing the normative work; the details are meant to show how optimizing for one task that looks fairly passive—i.e., retrieval—leads directly to active and generative procedures. That is, the model view of memory says that just by caring about retrieving relevant and accurate information, we end up with a memory system that

functions not like a file-box, but like an off-line deliberation system. It moves information around, builds theories, puts things together and fills in the gaps. The key result is not how the various stages work, but the fairly abstract fact that achieving the central epistemic goal of memory involves making new content. In a sense, the algorithmic level is just here to demonstrate the relationship between retrieval and active generation. This is where it is especially key that the model system is not a system for storing models, but a system for making them.

Put another way, if you are interested in the question of how we are justified in maintaining beliefs over time, you might think that the information-transforming aspects of memory are not relevant to answering this question. While it might be adaptive to come up with new information in memory, it need not effect my justification in believing based on memory in the “normal” case where the memory is not significantly altered. But I hope to have shown that this analysis which abstracts away from the messy cases is misguided. This approach misses the commonality between the “normal” case where a memory resembles a past belief closely, and other cases where the memory is accurate but has drifted significantly from sharing a content with any past belief. That is, both cases are the product of computations that enabled those contents to be retrieved, and both reflect background beliefs and values of the agent. Since computations like modeling are the sorts of things we are used to evaluating for rationality and justification, it should be at least an open question whether the etiology behind memory retrieval, even of “normal,” seemingly-unaltered content, is relevant to determining the justification of memorial beliefs.

7 | CONCLUSION

A feature of the model system is that it has a significant tendency towards incoherence; since we are building different kinds of models that idealize away from different features on the same data (i.e., they are overlapping), we should expect some of the extensions and extrapolations to contradict one another. Oddly enough, this shows a strength of the model view. That is, allowing memory to itself build and change models gives us a theory on which those changes are necessarily off-line (very off-line, in the case of human memory changes during sleep). Incoherence has worried a lot of philosophers, but the model view gives us a case of incoherence strictly speaking, but stripped of its most problematic features. This type of incoherence is between models, rather than within a model, and is processed at a distance from resultant actions (i.e., it is off-line).

A second consequence worth pointing out is that I have talked a lot about how the model system works, but my proposed change to the internal workings of memory also would lead to changes in the external role of memory, by which I mean the way other parts of our cognitive faculties are hooked up to memory. A rough way of characterizing this second shift is to think of how we use a set of facts from a file-box versus how we use a model. Where the former gets inputted and digested, the latter acts more like a guide.

In summary, in this paper, I emphasized the centrality of retrieval to memory, and situated the memory problem in a series of environmental conditions (a)–(c). From this starting point, I considered how memory changes during sleep give us a reason to look for an alternative epistemology of memory that explains why patterns and generalizations over stored information are enhanced and sometimes learned in sleep. It is these structures that explain the more commonly cited cases of constructive memory. I then presented two kinds of system that solve the retrieval problem in part by structuring and generalizing. My preferred option, the model system, makes, updates and maintains a series of overlapping representations that are simplified, domain-specific and relational. I argued that for realistic human creatures as well as ideal, non-limited agents, the model system works the best,

and the alternative index system, succeeds insofar as it starts to resemble the model system. In the final sections, I contrasted this essentially productive memory system to the systems implied by two families of theories in the epistemology of memory. This might have convinced you that these other theories need some work, but I was mainly hoping to demonstrate that adopting the model view will result in non-trivial changes to the epistemic role of memory.

I will leave the reader with this thought: philosophers have long studied on-line, deliberative changes in belief. But it is evident that many significant changes happen offline, whether transitions from believing P to believing $\neg P$, or realizing that Q is also a possibility, or so on. I suspect that looking closer at the normativity of information transformation in the memory system will be a significant part of this broader project of the epistemology of off-line belief formation and change.

ACKNOWLEDGEMENTS

I am grateful to Elisabeth Camp, Megan Feeney, Reza Hadisi, James Joyce, Sydney Keough, Victor Kumar, Richard Lewis, Peter Railton, Susanna Schellenberg, Eli Shupe, Chandra Sripada, Brian Weatherson and Timothy Williamson for invaluable comments.

REFERENCES

- Anderson, J. R. & Milson, R. (1989). Human memory: An adaptive perspective. *Psychological Review*, 96(4), 703–719.
- Aronowitz, S. (2018). Retrieval is central to the distinctive function of episodic memory. *Behavioral and Brain Sciences*, 41, E2.
- Barnett, D. J. (2015). Is memory merely testimony from one's former self? *Philosophical Review*, 124(3), 353–392.
- Bendor, D. & Wilson, M. A. (2012). Biasing the content of hippocampal replay during sleep. *Nature Neuroscience*, 15(10), 1439–1444.
- Bernecker, S. (2010). *Memory: A philosophical study*. Oxford: Oxford University Press.
- Burge, T. (1997). Interlocution, perception, and memory. *Philosophical Studies*, 86(1), 21–47.
- Buzsáki, G. & Moser, E. I. (2013). Memory, navigation and theta rhythm in the hippocampal-entorhinal system. *Nature Neuroscience*, 16(2), 130–138.
- Campbell, J. (2002). *Reference and consciousness Oxford cognitive science series* (). Oxford: Clarendon Press.
- Chalmers, D. J. (2004). The representational character of experience. In B. Leiter (Ed.), *The future for philosophy* (pp. 153–181). Oxford: Oxford University Press.
- Cicero, M. T. (55 bce/1948). *Cicero: On the orator, Books I-II (Loeb Classical Library No. 348)* (English and Latin Edition), trans. Sutton, E. W. & Rackham, H. Cambridge: Harvard University Press.
- Colgin, L. L., Moser, E. I. & Moser, M.-B. (2008). Understanding memory through hippocampal remapping. *Trends in Neurosciences*, 31(9), 469–477.
- Dalgleish, T., Navrady, L., Bird, E., Hill, E., Dunn, B. D. & Golden, A.-M. (2013). Method-of-loci as a mnemonic device to facilitate access to self-affirming personal memories for individuals with depression. *Clinical Psychological Science*, 1(2), 156–162.
- Diekelmann, S., Wilhelm, I. & Born, J. (2009). The whats and whens of sleep-dependent memory consolidation. *Sleep Medicine Reviews*, 13(5), 309–321.
- Dissel, S., Melnattur, K. & Shaw, P. J. (2015). Sleep, performance, and memory in flies. *Current Sleep Medicine Reports*, 1(1), 47–54.
- Donlea, J. M., Thimman, M. S., Suzuki, Y., Gottschalk, L. & Shaw, P. J. (2011). Inducing sleep by remote control facilitates memory consolidation in drosophila. *Science*, 332(6037), 1571–1576.
- Drosopoulos, S., Windau, E., Wagner, U. & Born, J. (2007). Sleep enforces the temporal order in memory. *PLoS ONE*, 2(4), e376.
- Fredkin, E. (1960). Trie memory. *Communications of the ACM*, 3(9), 490–499.
- Griessenberger, H., Hoedlmoser, K., Heib, D., Lechinger, J., Klimesch, W. & Schabus, M. (2012). Consolidation of temporal order in episodic memories. *Biological Psychology*, 91(1), 150–155.
- Gross, A. L., Brandt, J., Bandeen-Roche, K., Carlson, M. C., Stuart, E. A., Marsiske, M. & Rebok, G. W. (2014). Do older adults use the method of loci? Results from the active study. *Experimental Aging Research*, 40(2), 140–163.
- Hemmer, P. & Steyvers, M. (2009). A bayesian account of reconstructive memory. *Topics in Cognitive Science*, 1(1), 189–202.
- Hitchcock, C. & Sober, E. (2004). Prediction versus accommodation and the risk of overfitting. *The British Journal for the Philosophy of Science*, 55, 1–34.
- Liu, T.-Y. (2009). Learning to rank for information retrieval. *Foundations and Trends in Information Retrieval*, 3(3), 225–331.
- Loftus, E. F. (1996). *Eyewitness testimony*. Cambridge, MA: Harvard University Press.
- Luria, A. R. (1968). *The mind of a mnemonist: a little book about a vast memory*. Cambridge, MA: Harvard University Press.
- Michaelian, K. (2011a). Generative memory. *Philosophical Psychology*, 24(3), 323–342.

- Michaelian, K. (2011b). The epistemology of forgetting. *Erkenntnis*, *74*(3), 399–424.
- Michaelian, K. (2016). *Mental time travel: Episodic memory and our knowledge of the personal past*. Cambridge, MA: MIT Press.
- Poe, G. R., Walsh, C. M. & Bjorness, T. E. (2010). Cognitive neuroscience of sleep. *Progress in Brain Research*, *185*, 1–19.
- Rauchs, G., Feyers, D., Landeau, B., Bastin, C., Luxen, A., Maquet, P. & Collette, F. (2011). Sleep contributes to the strengthening of some memories over others, depending on hippocampal activity at learning. *The Journal of Neuroscience*, *31*(7), 2563–2568.
- Sweatt, J. (2009). *Mechanisms of memory*. London: Elsevier Science.
- Tse, D., Langston, R. F., Kakeyama, M., Bethus, I., Spooner, P. A., Wood, E. R., ... Morris, R. G. (2007). Schemas and memory consolidation. *Science*, *316*(5821), 76–82.
- Walker, M. P. & Stickgold, R. (2004). Sleep-dependent learning and memory consolidation. *Neuron*, *44*(1), 121–133.
- Xia, Y., Cheng, R., Prabhakar, S., Lei, S. & Shah, R. (2008). Indexing continuously changing data with mean-variance tree. *International Journal of High Performance Computing and Networking*, *5*(4), 263–272.
- Zhang, W. & Luck, S. J. (2009). Sudden death and gradual decay in visual working memory. *Psychological Science*, *20*(4), 423–428.

How to cite this article: Aronowitz S. Memory is a modeling system. *Mind Lang.* 2019;34: 483–502. <https://doi.org/10.1111/mila.12220>