Simulationism and Memory Traces

Felipe De Brigard
Duke University

Abstract

In the philosophy of memory there is a tension between a preservationist and a constructivist view of memory reflected in the debate between *causalism* and *simulationism*. Causalism is not only committed to the claim that there must be an appropriate causal connection between the remembered event and the content represented at retrieval but also that such connection is possible because of a content-preserving memory trace. Simulationism, by contrast, rejects the need for an appropriate causal condition and, thereby, makes the appeal to memory traces unnecessary. In this paper I argue that while the are strong conceptual arguments and empirical evidence to support a constructivist view of memory, the empirical evidence also suggest that the initial formulation of simulationism needs to be revised. In particular, I argue, first, that simulationism's commitment to a single cognitive system for mental time travel is likely wrong, and second, that simulationism cannot get rid of memory traces altogether, as they are still explanatorily indispensable when it comes to explaining a large number of memory-related phenomena. At the end, I end up suggesting a way of thinking about memory traces that, I think, is compatible with simulationism and a constructivist view of memory. If the view I put forth is on the right track, then an updated version of simulationism can help to dispel the dichotomy between causalism and simulationism as a false one.

Key words

Simulationism, Causalism, Memory Traces, Default Mode Network, Hippocampus, Mental Simulation.

Simulationism and Memory Traces

1. Introduction

 I started graduate school almost twenty years ago, as a student in both philosophy as well as psychology and neuroscience. As a result, I was exposed to two very different views on the nature of memory. On the one hand, there was the philosopher's view, according to which the primary function of memory is to reproduce a past event or experience. Such reproduction—this view holds—is underwritten by the preservation of the encoded content in a memory trace, which is later retrieved in the act of remembering. Moreover, memory is thought to be essentially distinct from imagination, not only because remembering, unlike imagining, requires a causal connection to the remembered event, but also because the verb "remembering" is thought to be factive. That is, for one to truthfully express the proposition "S remembers that p", then it must be the case that p obtained. One cannot remember what didn't happen; at best, one can imagine it. Sure, one may experience from time to time mental contents that do not correspond to actual events in one's past, or that perhaps distort them a bit, but such cases of false and distorted memories are, according to the philosopher's view, instances in which memory malfunctions (Kurtzman, 1983).

 On the other hand, there was the psychologist's view of memory. According to this view, memory isn't reproductive but reconstructive. Memories are encoded, not as individual events, but as instances of larger knowledge structures one has acquired through life (Bartlett, 1932). As such, every act of encoding is embedded within an act of retrieval of related information that interprets and transforms the experienced content. There is also loss of information; because perception is fast and often incomplete, and because attention is limited and filtering, not everything we experience is encoded. And, of course, there is forgetting. Memory traces, if they exist at all, likely don't preserve all of the encoded content but, at best, an incomplete shadow of a past experience. Remembering is thus a reconstructive process in which past experiences are put back together by the joint operation of an incomplete memory trace and an active imagination that helps to fill the gaps at retrieval. It is because of this essential interaction between memory and imagination that false and distorted memories occasionally occur, although they do not reflect a failure in the system, but rather a natural by-product of the reconstructive operations of our mnemonic processes. "Remembering", thus, need not be factive, for it is natural to remember events that did not happen or that, when brought to mind, differ in subtle ways from how they actually occurred.

 This tension between the preservationist view of the philosopher and the reconstructivist view of the psychologist is reflected today in two of the leading theories in the philosophy of memory: *causalism* and *simulationism*. The views are thought to be diametrically opposed and, as it happens in philosophy, practitioners tend to take sides. I did, too, about a decade ago (De Brigard, 2014). Yet, time has passed, arguments and counterarguments have been published, and new—and old—empirical findings have surfaced. It behooves us then to act wisely and to proportion our beliefs to the available evidence (Hume, 1748/1977). And I think the evidence suggests that simulationism, in its original formulation, is wrong, and that it needs to be reformulated. Moreover, contra the original formulation of simulationism, I think we need to appeal to memory traces to explain remembering. The good news, though, is that reformulating simulationism will allow us to see the dichotomy between causalism and simulationism as a false one, opening thus the door for a reconciliation. To that end, I will start, in section 2, with a brief recount of both the motivations for, and standard formulations of, both causalism and simulationism. Next, in section 3, I will argue that two assumptions of the simulationist view are likely false and, thus, that the

theory needs to be updated. Finally, in section 4, I show how this updated version of simulationism can help to dispel the false dichotomy between causalism and simulationism.

2. Causalism and Simulationism

Although the idea that memories are causally linked to the past events they are about can be traced as far back as Aristotle (De Brigard, 2023), causalism in its contemporary form is typically associated to the celebrated paper *Remembering* by Martin and Deutscher (1966). I like to think of the original motivation behind causalism as threefold. First, there was a patent dissatisfaction with classical representationalist views of memory, as they tended to advocate for internal or subjective criteria to distinguish memories from imaginations. Both rationalist and empiricist philosophers held that memories were preserved ideas of past perceptions or experiences, and that they could be distinguished from imaginations thanks to some internal or subjective criterion, such as vivacity, familiarity, coherence with other beliefs, or even apperceptions (De Brigard, 2019). Such "memory markers", unfortunately, are problematic, as it was easy to come up with counterexamples showing that neither of them constitutes necessary nor sufficient conditions for a present mental representation to count as a memory, as opposed to a perception or an imagination. A second motivation stem from an increasing dissatisfaction with direct realism, the view according to which remembering involved no intermediate representations, but rather a particular kind of direct acquaintance with the remembered past event. Alas, the mysterious nature of this direct acquaintance relation with the past was metaphysically suspicious, and by the mid 20th century direct realism was all but discredited (Furlong, 1948). Finally, there was also discontent with the behaviorist alternative which, driven by Wittgenstein's criticisms against the use of mental representation to explain psychological phenomena, inspired Malcolm to reject the claim that a proper account of remembering required talk of memory traces or any causal connection with the remembered event at all (Malcolm, 1963).

However, as Martin and Deutscher (1966) neatly show, the need for a causal connection to the remembered event becomes indispensable when we try to distinguish cases of actual remembering from cases of apparent imagining, apparent remembering, and relearning (Robins, 2016; Michaelian and Robins, 2018). Cases of apparent imagining involve individuals that bring to mind a mental content they think they are conjuring up anew, but turn out to be recollections of past events whose acquisition they had forgotten about. Cases of apparent remembering involve individuals entertaining mental contents they take to be recollections but in fact correspond to no past experiences at all. And, finally, cases of relearning involve individuals that have encoded particular contents, then forget about them, and then relearn them from a different source or via some deviant causal chain. Thus, to avoid these kinds of cases, and secure the need for a causal condition, Martin and Deutscher (1966) offered a causal view of remembering, according to which an individual, *S*, remembers a past event, *e*, if and only if:

1) *S* now, at retrieval, represents *e* [*Current representation*]
2) *S* represented *e* at the time of encoding [*Past representation*], and
3) There is an appropriate causal connection between the content represented at encoding and the content represented at retrieval [*Appropriate causation*]

The qualifier "appropriate" here is critical for the view, as it allows it to rule out cases of remembering that occur due to deviant causal chains or serendipitous relearning. Moreover, the appropriateness of the causal connection is in turn safeguarded by the existence of a stored memory

trace, representing the content formed at encoding, and recovered, unchanged, at retrieval. Note that although there are some variations among causalists—some, for instance allow certain differences between the encoded and retrieved event (Bernecker, 2009)—all of them accept the necessity of the causal connection between the experienced and the remembered event.

However, in the past decade, the necessity of an appropriate casual connection between the experienced and the remembered event has been questioned by simulationism, a view motivated by two lines of empirical evidence. First, there is an overwhelming amount of evidence showing that remembering is often inaccurate, distorted and false. Likely many memories we take to be true, and with which we live our lives seamlessly, are imprecise or wrong. Yet, the evidence also shows that false and distorted memories aren't haphazard, but rather plausible and consistent with a person's acquired experience and the conditions of recall. To illustrate, consider a classic study by Brewer and Treyens (1981) in which participants were asked to wait in a regular academic office while the researchers worked on setting up the experiment. Unbeknownst to the participants, though, the waiting office itself was the experimental setting. Every object in the office was carefully placed, with some being consistent with the "schema" of an office (e.g., telephone), and others inconsistent (e.g., cowboy hat). Participants were asked to wait in the office for a little while, and then were taken to a different room for a surprise memory test. In it, participants were given a list of objects, and they were asked to remember which of them were in the office they were just at. The list included both office consistent and office inconsistent "old" objects—i.e., objects effectively present in the office—as well as "new" ones, some consistent and some inconsistent with what one would normally find in an academic office. Revealingly, the results showed that participants were more likely to endorse as old (i.e., "false alarm") new items that were consistent with the schema (i.e., "lures") relative to inconsistent ones.

Many other influential studies, including well-known manipulations such as the DRM paradigm (Roediger and McDermott, 1995)—which produces high false alarm rates to semantically associated word-lures in a study list—as well as the many variations of the "lost in a shopping mall" study (Loftus and Pickrell, 1995)—in which experimenters manage to generate memory experiences of entirely fabricated events that nonetheless are plausible and consistent with the participant's background knowledge and history (Garry et al., 1996)—convincingly demonstrate that false and distorted memories are common and have an air of plausibility or schema-consistency to them.

How can we then square a view of memory as reproductive and of remembering as necessarily linking the encoded content with the retrieved one, with the empirical fact that people often have recollective experiences of items or events they never experienced? Does this mean that our memory system is constantly malfunctioning? Or does it mean that true and false memories are produced by two independent processes? But, if so, why would these two processes be entirely opaque to the subject's awareness? Many researchers argue that the evidence from studies on false and distorted memories speaks against the view that memory is merely reproductive, and suggest instead a view according to which memory should be thought of as reconstructive (Schacter et al., 2000, De Brigard, 2014). According to this view, remembering is not the retrieval of a memory trace where the exact same content stored during encoding is brought back to mind but, instead, it involves the reconstruction of a mental representation aimed at depicting a past event, in a process that may or may not employ stored information acquired in a single past experience.

Now the question is: why would memory be reconstructive? The answer to this question comes, in fact, from the second line of research that has inspired simulationism: the discovery that the neural mechanisms required for episodic memory are also necessary for engaging in certain

kinds of imaginations or mental simulations. This line of research dates as far back as 1965, when a classic study on amnesia by Talland documented that patients with Korsakoff's amnesia were unable to think about their future. Twenty years later, Tulving (1985) described parallel difficulties in K.C., a famous amnesic case. This observation inspired Tulving to think of episodic memory as a capacity within a larger cognitive system for "mental time travel", thanks to which we are also able to engage in both episodic past and future thought. Indeed, in the last 25 years, the view that our capacity to episodically remember our past and imagine our future are profoundly connected has received substantial support from many scientific fields, including neuropsychology, cognitive neuroscience, and developmental and comparative psychology (Schacter et al., 2015). In particular, it has been consistently shown that mental time travel engages the brain's default mode network (DMN), a set of functionally connected brain regions involving the medial and dorsolateral prefrontal cortices, the posterior cingulate cortex, precuneus and inferior parietal lobule, and the lateral and medial temporal lobes, including the hippocampus (Bucker et al, 2008). More recently, it has also been shown that the DMN is engaged in other kinds of episodic simulations, such as perspective taking (Spreng and Andrews-Hanna, 2015) and episodic counterfactual thinking (i.e., our capacity to imagine alternative ways past events could have occurred but did not; De Brigard et al., 2013; De Brigard and Parikh, 2019). As a result, some philosophers (De Brigard, 2014; Michaelian, 2016) and some neuroscientists (Addis, 2020) have argued that the "DMN is the brain's event simulator" (Addis, 2018), and that remembering should be seen simply as a particular operation of this larger episodic simulation system.

Together, these two lines of evidence—one on false and distorted memories and one on the common mechanisms underlying several forms of episodic simulation—motivated some philosophers to reject casualism in favor of a constructivist account in which remembering is a particular instance of a more general capacity to mentally simulate hypothetical personal episodes. Perhaps the most precise articulation of simulationism comes from Michaelian (2022), according to which an individual, *S*, remembers a past event, *e*, if and only if:

1) *S* now represents *e* [*Current representation*] and
2) *S*'s current representation of *e* is produced by a properly functioning and hence reliable episodic construction system that aims to produce a representation of an event belonging to *S*'s personal past [*Proper function*].

Unlike causalism, then, simulationism rejects the need for a past representation condition and for an appropriate causation condition, suggesting instead that all we need is a properly functioning simulation system that can produce reliable representations of past personal events.

Thus characterized, simulationism owes us an explanation as to how exactly should we interpret the proper function condition. As I understand it, Michaelian's interpretation is along the lines of reliabilism in epistemology, so that the episodic construction system is reliable if and only if it consistently produces true memories (Michaelian, 2016). I find this response somewhat unsatisfactory, not only because it makes it a brute fact of the system that it is reliable without explaining how or why, but also because it renders the proper function condition a target of well-known arguments against reliabilism in epistemology (Goldman and Beddor, 2021). One such challenge is known as the "generality problem" (Feldman, 1985), and the shape it adopts for the context of memory is that of determining, for any given instance of remembering, which memory forming cognitive process is responsible for its being true. More precisely, the worry is that there

is no principled way of individuating the cognitive process of type-remembering such that it could tell us, for a particular token-remembering, whether or not it has been reliably produced.

Lyons (2019) offered a solution to the generality problem in epistemology that is consistent with my own take on the proper function condition for remembering. Essentially, his suggestion is that we can go from type-general to token-specific process if we think of each belief—or, in our case, each memory—as produced by a computational algorithm that is constrained by different parameters. If a particular instance of remembering is generated by a process whereby the values taken up by each variable are within the range of the relevant computational parameters, then that memory has been reliably produced. In other words, the solution for the generality problem is to understand the relevant cognitive process in computational terms. Likewise, in 2014, my own take of the proper function condition in simulationism was to think of the process of memory reconstruction as carried out by a series of computational processes aimed at outputting the optimal solution given their current input.

The devil is in the details, of course, and back then I only glossed over the computational architecture of the reconstructive processes carried by the simulation system (De Brigard, 2012; 2014). At that point, the view I endorsed was inspired by the so-called "rational analysis" of memory (Anderson, 1989; for a recent review, see Gershman, 2022). According to this approach, cognitive processes can only be understood when considered as adaptations to their environments. In the case of memory, we know for instance that our environment is relentlessly bombarding us with more information than we can perceive, that a lot of the information we manage to perceive won't get stored, and that a lot of the information that gets stored will decay and be forgotten overtime. The data from which we have to reconstruct our memories is thus noisy and incomplete. We also know, though, that there are all sorts of statistical dependencies in our environments: some events are followed by others with high frequency, while others are rather rare or their occurrence is stochastic. The task, then, is to understand how such a computationally limited cognitive system can exploit those statistical dependencies to solve an informational retrieval problem in a way that is adaptive for organisms like us. Moreover, according to the rational analysis, the adaptive purpose of memory is future-oriented, that is, we need to retrieve accurate information about the past for future purposes. Unfortunately, the process of retrieval is costly. As such, the optimal computational strategy is going to be one that maximizes the odds of a gain (i.e., a successful retrieval of an accurate memory) while minimizing the costs (i.e., failing to retrieve an accurate memory or retrieving an inaccurate one).

This computational framework has the advantage of seeing remembering as a computational process whose purpose is *not* to faithfully preserve and/or reproduce a past content, but to optimally infer from an effect (i.e., the current representation) its most likely cause (i.e., the past representation). In other words, the computational problem our memory system is trying to solve is a variant of what is known as "an inverse problem": the challenge of determining, given a particular effect, what its cause must have been. Given the noisy and incomplete nature of the information the reconstructive process starts off with, the result is going to heavily rely on completions that are highly probabilistic and dependent on background experience and conditions of recall. And now we see how a simulationist account of remembering, whereby memory retrieval is thought as an optimal probabilistic reconstruction of a past experience given both a current noisy content and prior constrains, can accommodate both aforementioned lines of evidence. First the engagement of common neural structures during episodic past, future and other episodic simulations occurs because they all recruit the same computational constructive processes. Second, the prevalence of schema-consistent false and distorted memories in everyday life is explained by

the fact that these very schematic constraints are responsible for the accurate reconstruction of a past experience. Most of the time the mental simulation constructed at retrieval is such that it accurately represents the targeted past event, but sometimes it does not. Nevertheless, in both cases the computational operations underlying the constructive process are identical, and in both cases the system is doing what it is supposed to do.

3. A change of heart

Simulationism can then do away with any need to include a causal claim in their account of remembering, as it does not make it necessary for an accurate memory to include as its content the very same information that the subject experienced in the past and is now remembering. A genuine memory could just as well be produced by the same computational processes without the need to include information directly caused by the original event. Moreover, simulationism can also remove the need to postulate memory traces, understood as preserved stand-ins for the encoded content, poised to be recovered at the time of retrieval. If an accurate memory is fully reconstructed at retrieval, talk about memory traces may become unnecessary.

However, in the decade since I published my own version of simulationism (De Brigard 2014), it has become evident that there are a number of empirical and conceptual issues that put pressure against some core claims I defended back then—although it is likely that these criticisms affect other versions of simulationism as well, such as Michaelian's (2016). What I seek to do in this section is to offer both empirical and conceptual reasons against two such core claims, namely that the brain's DMN is an episodic construction system (it isn't), and that we don't need memory traces to explain remembering (we do).

3.1. Simulationism 2.0

According to simulationism, remembering is produced by a single cognitive system whose function is to generate episodic simulations, with memories being but a subset of them. Such an episodic construction system, the view goes, corresponds to the brain's DMN. After all, thinking of the DMN as the neural structure subserving episodic simulation helps to explain the commonalities between episodic recollection and other kinds of episodic simulations that emerge in parallel during development and are equally affected in individuals with brain damage. For instance, individuals with hippocampal damage have difficulty generating episodic memories and also episodic future and counterfactual thoughts (Schacter et al, 2015). Likewise, individuals with damage in the medial prefrontal cortex have difficulty spontaneously retrieving episodic autobiographical memories (Belfi et al, 2018) as well as generating episodic hypothetical thoughts (Beldarrain et al., 2005). Indeed, the idea that self-projection or self-simulation constituted a unified psychological kind, became an attractive theory to explain the function of the DMN when it was initially functionally characterized (Carroll and Buckner, 2007).

Unfortunately, there is quite a bit of counterevidence that is hard to accommodate within this view. First of all, there are now many experimental results showing that activity in the DMN is associated with all sorts of cognitive functions that are hard to fit under the umbrella of "episodic simulation", including semantic processing (Lanzoni et al, 2020), allostasis and interception (Kleckner et al, 2017), addiction (Zhang and Volkow, 2019), and aesthetic appeal (Belfi et al., 2019), among others. It would be a stretch, I think, to try to group all of these cognitive functions under the same general category of "episodic simulation". Alternatively, if everything counts as "episodic simulation", then the explanatory advantage of the construction system in explaining remembering would be severely diminished.

A second concern is that many non-human animals have a DMN, yet it is very questionable whether they can engage in complex episodic simulation. Neural evidence has revealed the presence of DMN in the brain of mice (Sforazzini et al, 2014), rats (Lu et al., 2012), rabbits (Schroeder et al., 2016), marmosets (Liu et al., 2019), and monkeys (Vincent et al, 2007). The presence of this brain network across such vastly different mammalian species suggests that it may not have evolved for complex cognitive functions such as the generation of episodic counterfactual thoughts or autobiographical recollections. A more parsimonious explanation is perhaps that the DMN reflects more basic metabolic and homeostatic processes, as initially hypothesized by Raichle and colleagues (2001), rather than constituting a unified cognitive system whose function is best characterized in psychological terms.

A third issue with the claim that the DMN is the simulation system of the brain is the fact that not all episodic constructive processes depend to the same extent of core regions of the DMN—especially the hippocampus. As it turns out, the engagement of the DMN in the construction of episodic simulations is highly dependent on the particular contents that are simulated. Research on episodic counterfactual thinking, for instance, has shown that thinking about alternative ways in which events could have unfolded in one's past recruits the DMN, but not so when it comes to thinking about what an unfamiliar person in an unfamiliar situation would have done (De Brigard et al., 2015). Likewise, the DMN is *not* preferentially recruited when it comes to thinking about episodic counterfactual situations that involve objects, as opposed to people (Parikh et al., 2018). Similar considerations apply to episodic simulations involving theory of mind, as patients with hippocampal damage can engage in such simulations as long as they do not demand retrieving information from episodic memory (Rosenbaum et al., 2007).

A fourth issue concerns navigation. A critical function of the DMN and, in particular, the hippocampus, is the capacity to mentally generate spatial simulations (Spreng et al, 2009). But even this association, I surmise, is questionable. For years, it's been thought that medial temporal lobe structures—mainly the hippocampus and entorhinal cortex—are both necessary and sufficient for navigation. However, in a recent study, Long and Zhang (2021) demonstrated spatial mappings in the somatosensory cortex of foraging rats, and shortly after Wikenheiser and colleagues (2021) found spatially-localized firing in neurons in the rat's orbitofrontal cortex that mimic those found in the hippocampus—the only difference being that they were less temporally precise. Critically, both the somatosensory and the orbitofrontal cortices are not part of the canonical DMN. Moreover, evidence from individuals with developmental as well as adult-onset amnesia reveals several preserved spatial navigation abilities despite their episodic memories being impaired (Rosenbaum et al., 2000; Rosenbaum et al, 2015). In fact, a very recent study of a patient with severe bilateral medial temporal lobe damage shows comparable performance on spatial navigation and spatial memory tasks despite abysmal results in autobiographical and episodic memory tests (McAvan et al., 2022).

Fifth, and perhaps more relevant for our current purposes, is the fact that individuals with hippocampal damage can still generate spatial simulations. One of the leading theories seeking to explain the common engagement of the hippocampus during episodic past, future, and other hypothetical simulations, has been defended by Maguire and colleagues (e.g., Maguire and Mullaly, 2013). According to this view, the hippocampus is required for us to be able to engage in the mental simulation of scenes in space. The problem is that even their own data suggests that patients with hippocampal damage can still think in and about space. In their landmark paper, Hassabis et al. (2007) showed that patients with hippocampal damage had difficulty thinking about new experiences, and they interpret such deficits as reflecting their incapacity to mentally simulate

spatial scenes. But take a look at the transcript of one of the (available) narratives from one of the patients, who is asked to imagine standing in the main hall of a museum:

> *Interviewer*: So, what does it look like in your imagined scene?
> *Patient P05*: Well, there is big doors. The openings would be high, so the doors would be very big with brass handles, the ceiling would be made of glass, so there's plenty of light coming through. Huge room, exit on either side of the room, there is a pathway and map through the centre, and on either side there'd be the exhibits.

How can one read this transcription and not think that this patient is mentally simulating a scene that occurs in space? How can one imagine that some things are on one side of a room, or that light comes through the ceiling, or that there is a map in the center of the room, if one can't envision spatial scenes? It is certainly the case that, when compared with controls, these mental simulations are less rich and detailed—a point I'll discuss below—but it would be false to say that they aren't reflecting an imagined spatial scene. Moreover, there is evidence that even H.M. was able to not only remember spaces from his old house, but he was also capable of mentally navigating through such spaces (Corkin, 2013).

And finally, contrary to some interpretations of the mental time travel system, it turns out there is also evidence to the effect that the hippocampus may not be required for thinking about time. Recent studies on K.C., the famous amnesic patient that motivated Tulving to talk about mental time travel in the first place, have shown that he has no problem talking about temporal concepts such as time travel, why events that happened in the past can't be modified in the present, why the future does not affect the past, or why people wouldn't do things now because they may regret them in the future (Craver et al, 2014a). K.C. is also subject to decision-making biases that allegedly require the capacity to anticipate the future, such as the Allais paradox (Craver et al., 2014b)—the tendency to give inconsistent answers in violation of expected utility theory when forced to choose between two gambits—and delayed discounting in intertemporal choice (Kwan et al, 2015). Neuroimaging evidence supports these findings, showing neural differences in brain activity associated with episodic future thinking versus delayed discounting in economic choices (Benoit et al, 2011).

Taken together, these various findings strongly suggest that the DMN is *not* a distinct cognitive system for the mental construction of episodic simulations. On the contrary, the evidence indicates that it is neither necessary—for it is possible to generate the kinds of contents such a system is supposed to produce without the engagement of core regions of the DMN—nor sufficient for episodic simulation, as some kinds of mental constructions of imaginative scenes recruit regions outside of the DMN.[1] In my view, what the available evidence supports is a much narrower claim: namely, than a healthy hippocampus—and hippocampal path (e.g., Ayala et al., 2022)—is needed to successfully engage in episodic recollection and certain kinds of episodic simulations. The question now is: what do these episodic simulations have in common?

A possible answer comes from thinking about the common engagement of the hippocampus during these kinds of episodic simulations as having less to do with *what* they

---

[1] A reviewer suggested that there is evidence to the effect that the DMN actually consists of at least three sub-networks (Andrews-Hanna et al., 2010), and that it may be possible that while the DMN as a whole isn't necessary or sufficient for episodic simulation, perhaps one such sub-region may be. While it may be possible that some parcellation of the DMN has a better chance at being identified as "the simulation system of the brain" (although I am very skeptical), my target here is the original formulation of simulationism, which was not confined to a sub-network of the DMN.

represent—i.e., what their contents are about—and more with *how* they represent it, i.e., the representational format or structure. The suggestion I put forth (and which I more fully articulated in De Brigard and Gessell, 2016), is that the kinds of episodic simulations relevant for understanding the neural substrates of mental time travel can vary along two dimensions: content and structure. Now, contents can be more or less about time; some of our thoughts are clearly about past or future events, some are clearly not, and some are in between. Call this dimension *tense*. But the structure of our thoughts can also vary along a temporal dimension: some thoughts take little time to be entertained and they don't seem to require much of a temporal structure to unfold; others, by contrast, need more time as their contents involve structures that need time to unravel. Call this dimension *dynamicity*. Remembering that Caracas is the capital of Venezuela, or what an apple looks like, are relatively static in that they don't require much time unfolding in working memory to be entertained, whereas remembering one's first kiss, imagining how things may have unfolded had one not missed a particular elevator ride, or thinking about fixing the dishwasher in the weekend, presumably require more. The more dynamic the structure of the mental simulation, I surmise, the more the hippocampus is needed.

As reviewed above, there is evidence suggesting that core regions of the DMN—and, in particular, the hippocampus—are neither necessary nor sufficient to engage in thoughts *about* space or time. But if we think of the hippocampus—at least in the human case—as required for episodic simulations that unfold *over* time, then the counterevidence is no longer threatening. Most of the tensed thoughts K.C. is capable of tend to be short and somewhat telegraphic, not unlike the reports from the amnesic patients in the Hassabis et al., (2007) study. For instance, patient P03, when asked to imagine "lying on a white sandy beach in a beautiful tropical bay," simply responds "all I can see is the colour of the blue sky and the white sand". Further probing elicits no more that "like I'm kind of floating". Similar to the transcript of patient P05 above, what these two narratives seem to have in common is that they lack dynamicity; it is as if they were describing a static picture in space, rather than immersing themselves in an episodic simulation that's unfolding over time in their minds. Indeed, additional evidence shows that individuals with hippocampal damage can describe spatial scenes, as long as they are static (Gaesser et al., 2010; Race et al., 2011; Race et al., 2013), as well as non-tensed fictional events that do not require much dynamicity or elaboration (Rosenbaum et al., 2009; Romero and Moscovitch, 2012).

This dynamic interpretation of the role of the hippocampus in episodic simulation also helps to account for some classic neuroimaging results in the mental time travel literature. For instance, in their landmark paper, Szpunar et al. (2007) found no difference in hippocampal engagement between their experimental conditions (i.e., episodic past and future thinking) and their control condition, whereas Addis et al., (2007) did. The reason, I surmise, is because in the former, the control condition was an episodic simulation involving a familiar other (i.e., Bill Clinton), whereas in the latter, the control condition was a sentence construction task, which does not require the dynamic deployment of a complex mental simulation. The recruitment of the hippocampus during non-mental time travel tasks, such as imagining fictitious (Hassabis et al, 2007) and non- temporal events (D'Argembeau et al., 2008), as well as possible personal past (Addis et al., 2009) and counterfactual episodes (De Brigard et al., 2013; van Hoeck et al., 2013), can also be explained by the fact that such simulations involve the generation of complex dynamic representations. And, incidentally, thinking about the role of the hippocampus in the generation of dynamic, as opposed to static, simulations, may help to resolve the theoretical conflict between the *scene construction* (Hassabis & McGuire, 2007) and the *constructive episodic simulation* (Schacter & Addis, 2007) hypotheses. Scene construction is defined as "the process of mentally

generating and maintaining a complex and coherent scene or event" (Hassabis and McGuire, 2007: 299). The view of the hippocampal role in episodic simulation I advocate for here is entirely compatible with the scene construction perspective insofar as the scene that is mentally generated and maintained is dynamic, structured and unfolds over time. This makes the scene construction view a hypothesis about the structure or the format of the mental simulation. By contrast, the episodic simulation hypothesis speaks to the etiology of the contents of our episodic simulation— whether they are drawn from episodic memories, or semantic and conceptual knowledge—not about their format. Both views are, then, compatible.

Finally, it is worth noting that although the hippocampus is traditionally associated with spatial representations, thanks to the discovery of place and grid cells, there is also quite a bit of evidence showing that these cells are also sensitive to sequential information about spatial navigation. Landmark studies by O'Keefe and Recce (1993) and Skaggs and colleagues (1996) showed that the moment of the firing of a place cell within a navigational sequence has a precise timing relation with oscillations in the theta band. Likewise, Foster and Wilson (2007) demonstrated that place cells in CA1 are "timed- locked" to theta oscillations, suggesting that, prior to performing a learned sequence, place cells can "pre-play" the upcoming action by firing in succession. These results suggest that the hippocampus, and the entorhinal cortex as well, are not only encoding the spatial but also the temporal relations among the components of a scene. The fact that the components of a simulated episode are bounded by temporal relations whose mental reinstatement requires time to unfurl may be as essential to the engagement of the hippocampus as the fact that the elements of the simulation stand in spatial relations too.

In sum, the purpose of this section has been twofold. First, I offered some counterevidence against the simulationist's claim that the DMN is the "episodic simulation system of the brain" (De Brigard, 2014; Addis, 2018; 2020). I argued, instead, that although the evidence does not point toward there being a single unified system for episodic simulation, it does suggest that the hippocampus—and likely adjacent regions in the medial temporal lobe—are required for the proper construction and maintenance of certain kinds of episodic mental simulations. What these kinds of episodic mental simulations have in common—and this is the second point I sought to put forth in this section—is that they are dynamic, in the sense that their contents are not only spatially but also temporally structured in a way that the very unfolding of the episodic simulation (i.e., its generation and maintenance) is protracted and takes time. Thus, simulationism needs to abandon the idea that there is a unified cognitive system for episodic simulation and, instead, adopt the narrower view that the commonality in hippocampal engagement has to do with the dynamic format of the episodic simulation (De Brigard and Gessell, 2016).

3.2. Probabilistic dispositionalism

Unlike causalism, simulationism sees no need for the appropriate causation condition. After all, the episodic construction system can output a reliable memory that neither represents nor was caused by an actual past event (Michaelian, 2022). This also means that simulationism has no use for memory traces. Recall that the notion of memory trace was introduced as a content-bearing theoretical entity that could help to bridge the causal gap between the encoded (i.e., past representation) and the remembered events (i.e., current representation). But if there is no prior representation, then there is no gap to be bridged, and thus there is no need for memory traces (Michaelian, 2016). Memory traces, therefore, appear to be incompatible with simulationism.

In this section, however, I argue against this claim. My argument has a negative and a positive part. In the negative part, I argue that, contrary to what simulationism holds, there are

many memory-related phenomena whose explanation makes it indispensable to appeal to memory traces. In the positive part, though, I argue that there is a promising account of the nature of memory traces that not only makes sense of why the hippocampus is involved in several kinds of episodic simulations (including remembering) but also why is it that false and distorted memories can be the normal product of a properly functioning reconstructive process. For reasons that I hope will become clearer in what follows, I call this view on memory traces "probabilistic dispositionalism".

There are at least two arguments for thinking that memory traces are indispensable when it comes to explaining the psychological process of remembering. The first one, which I've developed at length before (De Brigard, 2020), has to do with differential effects in recollection and it is directed against a well-known argument against memory traces from Malcolm (1977). Many philosophers—including Malcolm—agree that memory traces are postulated via inference to the best explanation in order to avoid either the metaphysically dubious direct realist route or the equally problematic acceptance of causation-at-a-temporal distance between a non-existent past event (i.e., the rember*ed* event) and a current one (i.e., the remember*ing* event; Heil, 1978; Bernecker, 2008). Since the justification for their postulation is an inference to the best explanation, then it follows—according to Malcolm—that accounts of remembering that make use of the notion of memory trace should be better than those that do not. But then Malcolm suggests that when I give an account of why is it that I remember a particular event, say, a boat capsizing, all I need to do is refer to my prior experience of having witnessed the boat capsizing. At no point do I need to invoke, in addition to my having witnessed the event, the existence of some unobservable causal intermediary memory trace. To not multiply entities without necessity, and given that both accounts are equally good, then there is no reason to postulate the existence of memory traces.

The problem with this argument is that, in only looking at cases of successful recollection, Malcolm overlooks a large swath of memory related phenomena for which talk about intermediary memory traces becomes indispensable. For in addition to successful recollection, we often demand causal explanations for cases of *unsuccessful* recollection. Suppose that I witness a boat capsizing alongside my friend Andrea. The following day, I recall the event but Andrea doesn't. It seems as though some causal story is needed in order to explain why is it that I can remember the event whereas Andrea cannot. Or suppose that I can remember more details than she can, or that my memory is profoundly distorted relative to hers. In a word, cases of differential recollection under the same encoding conditions highlight the need to include some story about causal intermediaries that can explain the difference at retrieval. Appealing to memory traces—or, if you are unhappy with the term, some kind of causal intermediary between encoding and retrieval—becomes indispensable again.

The second argument is related, though it involves the use of a pharmacological agent—propranolol—during memory reactivation. One of the most interesting findings in the memory literature in the past few decades, is the fact that when a memory is reactivated at retrieval, it becomes labile and prone to modification (Nader & Einarsson, 2010). The evidence suggests, in fact, that there is a window of time in which, if the reactivated memory is intervened upon, its contents can be modified, even erased. A pharmacological intervention using propranolol, a synthetic beta-adrenergic receptor blocker that acts as an inhibitor of protein synthesis underlying memory consolidation and re-consolidation, has been shown to successfully extinguish both stimulus and context specific conditioning when administered immediately after memory reactivation and prior to reconsolidation (e.g., Leal Santos et al., 2021). Importantly, these effects

are condition-specific, meaning that the administration of propranolol only affects the retention of memories tied to a particular spatiotemporal context. Other memories that may have been acquired before or after, are unaffected. Given the specificity of these effects, it seems extremely difficult to explain such results without assuming that something in the brain of the animal was altered, and that such neural substrate is tied to a specific experienced event. Once again, the need to posit memory traces—or at least intermediate causal mechanisms—becomes evident.

The explanatory indispensability of a theoretical term is certainly not a sufficient condition to accept the existence of its putative referent. Nevertheless, in the case of memory traces, it has been a good motivator to try to discover them, and one that clearly has inspired neuroscientists for decades to understand what their nature might be (Josselyn and Tonegawa, 2020). My purpose, in the rest of this section, is to offer a general framework to characterize the nature of memory traces such that it can accommodate both the fact that memories are often distorted and false and the narrower thesis, defended in the previous section, that the hippocampus is required for the successful construction of dynamic episodic simulations, including episodic memories.

But first: some conceptual clarifications. Views on memory traces vary along several dimensions. On the one hand, some of those dimensions affect the *vehicle* of the mnemonic representation. For some researchers, memory traces are localized representations, akin to discrete symbolic entities carrying a particular mnemonic content, and instantiated in dedicated neuronal populations or even in specific cells (Gallistel, 2010). For others, memory traces are distributed across neuronal connections (Rumelhart and McClelland, 1986), and for some they are a combination of both (Gershman, 2022). Moreover, some even think of memory traces as extended and/or embodied, meaning that the vehicles carrying the relevant mnemonic content extend beyond the limits of the brain. On the other hand, some dimensions pertain to the degree of explicitness of the mnemonic content. Some hold that memory traces carry stripped-down versions of the encoded content but they do so explicitly, meaning that, in principle, one could read off the content directly from the neuronal substrate. For other, contents are implicitly encoded, for an additional process— i.e., retrieval—is needed to make them explicit. And, finally, some argue that contents are not occurrent—and, thus, are neither explicitly nor implicitly encoded—but rather dispositional, i.e., what gets encoded is a disposition to revive a content at retrieval given the right cue. Each one of these views has advantages and disadvantages, the discussion of which unfortunately I have to sidestep (but see De Brigard, 2014 and De Brigard, 2023). Nevertheless, I hope this brief excursus on the terminology is sufficiently informative, as the view I seek to defend in what follows is a version of dispositionalism predicated on representational vehicles, not for contents, whereby a memory trace is the disposition of a neural network to reactivate, as approximately as possible, the state it was in during encoding at the time of retrieval. Let's see how this works.

The framework I have in mind builds upon the hippocampal indexing theory (HIT), initially put forth by Teyler and DiScena (1986), in order to explain how memory traces formed during encoding could be reinstated at retrieval. Consistent with the complementary learning systems model (McClelland et al., 1995), HIT postulates that when an event is experienced, two consolidation processes take place. First, there is a fast *cellular* consolidation in which information is encoded as changes in connectivity among the neurons involved in the initial processing of the encoded event. Second, there is the more protracted *systems* consolidation, in which hippocampal-neocortical connections are further strengthened (Figure 1)[2]. As an illustration, consider how the

---

[2] When initially proposed, HIT followed the standard model of memory according to which, once consolidated, the hippocampus was no longer required for retrieval. However, further examination of extant data as well as new evidence indicates that the hippocampus is still required for the retrieval of recent and remote memories (Nadel and

model explains the formation of a memory trace for, say, the experience of seeing a boat capsizing. Walking by the shore you peer at the horizon and notice a vessel tipping over. This experience engages several regions of your cortex: sounds will be processed by the auditory cortex, shapes, colors and the like will be processed by the visual cortex, smells by the olfactory cortex, and so forth. Active neurons in the hippocampus form an *index* (more about this term in a second) that binds these distributed cortical patterns into a larger hippocampal-neocortical network which, over time, becomes systems-consolidated. Now suppose that a few days later, someone asks you if you've ever seen a boat capsizing. This auditory cue helps to re-activate a portion of the cortical pattern, and this reactivation propagates to the hippocampal index, which in turn enables the reactivation of the rest of the neuronal pattern, effectively reinstating the encoded hippocampal-neocortical network. Since the vehicle of the encoded representation is reactivated, then the encoded content is reenacted and, thus, you manage to remember the event.
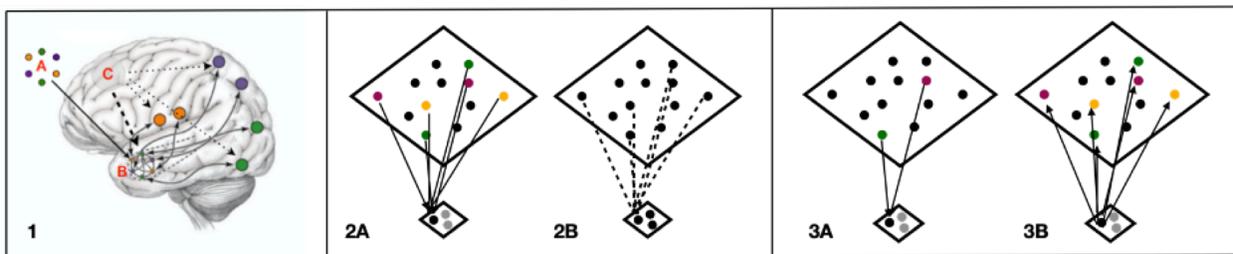
Figure 1. *Hippocampal Indexing Theory [HIT]. 1. Graphical schematic in brain space.* An initial stimulus with multiple sensory properties is first experienced (A). A rapid consolidation occurs in the hippocampus (B) while the sensory information of the stimulus is processed in the relevant areas of the sensory cortex (i.e., visual cortex for visual properties, auditory cortex for auditory properties, etc.). This co-activation creates an association between the sensory regions and a hippocampal index. At retrieval, a top-down signal from the pre-frontal cortex (C) to elements of the hippocampal-cortical assembly reactivates the network and, thus, the memory content. *2. Encoding.* The bigger layer indicates units in the neocortex, with different colors indicating different sensory information. The smaller layer indicates specific synaptic activity uniquely associated with the pattern of neocortical activation (2A). After encoding, consolidation strengthens the connection between the hippocampal index and the associated neurons in the neocortex (2B). *3. Retrieval.* A cue can reactivate a subset of the neocortical pattern, which in turn reactivates the hippocampal index (3A). This reactivation further spreads to the rest of the hippocampal-neocortical network, effectively reinstating the encoded pattern (3B). (Original figure from De Brigard, 2023).

The notion of *index* requires further clarification. As far as I know, the first reference to a hippocampal index is from Marr (1974), who called it a "simple memory", meaning a kind of sketchy or abstract representation of the encoded event. According to this view, what the hippocampus does is store a low-dimensional representation—perhaps akin to a lossy compression format (e.g., JPEG for images) or a set of principal components—out of which the encoded representation can be reconstructed. I tend to disagree with this view, for three reasons. The first reason is that compression formats require decompressing processes, and it is unclear what would be doing the decompressing in the hippocampal case. The second reason is that having a second representation that allegedly encodes more or less the same information as the first one, makes the hippocampal activity rather redundant. And, finally, it seems unlikely that all the neocortical variability that is required to represent different modalities and formats of information can be

---

Moscovitch, 1997). The version of HIT I advocate for here is consistent with this further development and has been further incorporated into a recent update of the CLS view (Kumaran, Hassabis and McClelland, 2016).

captured by the sparser archicortex of the much smaller structure that the hippocampus is. In fact, when trying to decode categorical and sensory information from brain activity at retrieval, multivariate patterns are unable to recover any encoded structure from hippocampal activity alone (Huffman and Stark, 2014). The evidence, therefore, does not suggest that the hippocampus is in fact storing a "simple memory".

Instead, I suggest that what the hippocampal index encodes is not an explicit—if compressed—content but rather a set of conditional instructions or dispositions to reactivate, as best as it can, the cortical pattern of activity it was associated with during encoding. Thus, when one experiences a certain event during encoding, the experienced content is instantiated in a particular representational vehicle, in the form of a hippocampal-neocortical network in the brain. Consolidation increases the probability of the nodes in the network to coactivate given the right cue. When such a cue is presented in the retrieval context, the coactivation among units of the network starts to propagate toward the hippocampal index, which does not contain explicit contents but rather the conditional instructions to reactivate the rest of the pattern of activity[3]. This, I suggest, is the right way to understand what a memory trace is: the dispositional property of a neural network to reinstate the state it was in, during encoding, at the time of retrieval. In fact, by characterizing the memory trace as a dispositional property of a representation vehicle rather than a content, one not only avoids concerns about content dispositionalism (Vosgerau, 2010), but also can readily explain why unexpected cues can bring to mind involuntary memories (Berntsen, 2009). Additionally, this view helps to explain why, when a memory is reactivated, it becomes modifiable. Since every act of retrieval is itself an act of re-encoding, nodes that weren't part of the original pattern but that already have a higher baseline probability of being coactive, are now more likely to getting included in the original pattern of activation after reconsolidation. Finally, this view also accommodates the fact that information acquired after the encoding and prior to retrieval can influence the way we remember past events. Here's an example I like to use. Long ago, before I learned how to speak English, I committed to memory the chorus of "A hard days' night" by the Beatles. I did not know what it meant, but I could sing the words. Years later, after learning English, I found myself listening to the song again, and was able to remember the lyrics and sing along. But as I was remembering the words, I was also understanding them for the very first time. The content of my recollection was different from that of its encoding, due to an intervening change to the network units that formed the representational vehicle of my memory.

4. Rapprochement

In the previous section I sought to defend two claims. First, contra simulationism, I argued that memory traces are often explanatorily indispensable to account for certain instances of remembering and memory related phenomena. Second, I offered a general framework, inspired by the HIT, as to how to understand memory traces and the role the hippocampus plays at retrieval. Specifically, I suggested that memory traces could be understood as dispositional properties of hippocampal-neocortical networks to reinstate the pattern of activation they were in, during encoding, at the time of retrieval. Since this pattern of activation consists in a set of nodes whose probability of coactivation is high—perhaps relative to some threshold—the reactivation is going to be sometimes imprecise and noisy.

---

[3] One can think about the end-state of the reactivation in terms of a Hopfield network (Hopfield, 1982), a kind of recurrent artificial network that tend to stabilize in a particular pattern of activation or "attractor". The idea, then, is that the final state of reactivation of the hippocampal-neocortical network is the attractor state of the network that underlies the encoded content.

Thinking about memory traces in this way—call this approach "dispositional probabilism"[4]—has the advantage of accommodating the two main motivations for simulationism. First, the fact that false and distorted memories are schema-consistent is a natural consequence of the probabilistic pattern completion process. As mentioned, a lot of the information that is initially perceived won't be encoded, and retention may in turn degrade some of the connection weights between units. But thanks to the statistical regularities in the connections of such units, the probability of reactivating the right set of connections given a cue remains high. This is the sense in which reconstruction can be said to be both probabilistic and veridical. However, since units in the hippocampal-neocortical network have additional existing associations with other units, it is possible that a unit or a set of units not involved in the encoding of the original content can become active during pattern completion at retrieval, resulting in a reconstructed memory that does not accurately represent the past event[5].

Second, dispositional probabilism can also explain why the hippocampus is engaged in certain kinds of episodic simulation, besides remembering. If we think of the hippocampal index as triggering the reactivation of different sensory areas in order to reenact the contents they process, and such reactivation takes time to complete into a single coherent scene, then we can think of this computation in analogous terms for episodic memory as well as for other dynamic mental simulation. More precisely, since the hippocampal index per se does not represent the information processed in the neocortex but rather rules to reactivate a more complex set of cortical units responsible for actually carrying out the contents of the mental simulations, the same reactivating computations can be recruited to recreate dynamic scenes that do not correspond to actual past but to hypothetical episodes. As such, it is not surprising that similar hippocampal-neocortical networks are recruited during certain kinds of episodic simulations because the computations that underlie their generation are not that dissimilar from those involved in episodic recollection.

Although there is much more that can be said about the computational nature of memory traces in general, and about my favored probabilistic-dispositional account in particular, I hope that what I've said enable us to see how we may be able to dissolve the conflict between causalism and simulationism. For simulationism need not postulate the existence of a dedicated system for episodic simulation, but simply accept that certain structures—particularly the hippocampus—are shared among certain kinds of episodic simulations. Such communalities are explained because the same underlying computations for dynamic pattern completion can be redeployed in certain kinds of episodic simulation. Additionally, simulationism can accept the existence of memory traces, as long as they are understood as dispositional properties of hippocampal-neocortical networks (i.e., representational vehicles) to probabilistically reenact the state they were in, during encoding, at the time of retrieval. Now, whether probabilistic dispositionalism can help to understand our tendency to think of remembering in causal terms is, alas, a story I need to leave for another day.[6]

---

[4] For similar views, see Vosgerau, 2010; Perrin, 2021; Werning, 2020.

[5] There is quite a bit of research in computational psychology and neuroscience trying to understand the precise computations that best characterize the probabilistic process of pattern completion at retrieval. In my opinion, a promising avenue is the "rational analysis approach" (Anderson, 1989) I mentioned before. On this perspective, extant associations between units can influence the pattern of neural activation by combining values reflecting prior frequencies as well as previously acquired conceptual/semantic associations.

[6] Previous versions of this paper were presented at the conference on *Simulationism* at the University of Grenoble, Alpes, in July 2022, the workshop on *Memory, Space and Time* at the University of Arizona, in November 2022, and the *Generative Episodic Memory* conference at Bochum, in June 2023. Many thanks to the organizers and the audiences of these events.

References

Addis, D. R., Pan, L., Vu, M. A., Laiser, N. and & Schacter, D. L. (2009). Constructive episodic simulation of the future and the past: Distinct subsystems of a core brain network mediate imagining and remembering. Neuropsychologia, 47, 2222-2238.

Addis, D.R. (2018). Are Episodic Memories Special? On the Sameness of Remembered and Imagined Event Simulation. Journal of the Royal Society of New Zealand, 48, 64-88.

Addis, D.R. (2020). Mental time travel? A neurocognitive model of event simulation. Review of Philosophy and Psychology (Special Issue on Mental Time Travel), 11, 233–259.

Anderson, J. R. (1990). *The adaptive character of thought*. Hillsdale, NJ: Erlbaum.

Ayala, O. D., Banta, D., Hovhannisyan, M., Duarte, L., Lozano, A., García, J. R., ... & De Brigard, F. (2022). Episodic past, future, and counterfactual thinking in relapsing-remitting multiple sclerosis. *NeuroImage: Clinical*, *34*, 103033.

Bartlett, F.C. (1932). *Remembering: A study in experimental and social psychology*. Cambridge university press.

Beldarrain, M. G., Garcia-Monco, J. C., Astigarraga, E., Gonzalez, A., & Grafman, J. (2005). Only spontaneous counterfactual thinking is impaired in patients with prefrontal cortex lesions. *Cognitive Brain Research*, *24*(3), 723-726.

Belfi, A. M., Karlan, B., & Tranel, D. (2018). Damage to the medial prefrontal cortex impairs music-evoked autobiographical memories. *Psychomusicology: Music, Mind, and Brain*, *28*(4), 201.

Belfi, A. M., Vessel, E. A., Brielmann, A., Isik, A. I., Chatterjee, A., Leder, H., ... & Starr, G. G. (2019). Dynamics of aesthetic experience are reflected in the default-mode network. NeuroImage, 188, 584-597.

Benoit, R. G., Gilbert, S. J., & Burgess, P. W. (2011). A neural mechanism mediating the impact of episodic prospection on farsighted decisions. *Journal of Neuroscience*, *31*(18), 6771-6779.

Bernecker, S. (2008). *The metaphysics of memory* (Vol. 111). Springer Science & Business Media.

Bernecker, S. (2009). *Memory: A philosophical study*. OUP Oxford.

Berntsen, D. (2009). *Involuntary Autobiographical Memories*. Cambridge University Press.

Brewer, W. F., & Treyens, J. C. (1981). Role of schemata in memory for places. *Cognitive psychology*, 13(2), 207-230.

Buckner, R. L., Andrews-Hanna, J. R., & Schacter, D. L. (2008). The brain's default network: anatomy, function, and relevance to disease. *Annals of the new York Academy of Sciences*, *1124*(1), 1-38.

Buckner, R. L., & Carroll, D. C. (2007). Self-projection and the brain. *Trends in cognitive sciences*, *11*(2), 49-57.

Corkin, S. (2013). *Permanent present tense: The unforgettable life of the amnesic patient H. M*. Basic Books.

Craver, C. F., Kwan, D., Steindam, C., & Rosenbaum, R. S. (2014). Individuals with episodic amnesia are not stuck in time. *Neuropsychologia*, *57*, 191-195.

Craver, C. F., Cova, F., Green, L., Myerson, J., Rosenbaum, R. S., Kwan, D., & Bourgeois-Gironde, S. (2014). An Allais paradox without mental time travel. *Hippocampus*, *24*(11), 1375-1380.

D'Argembeau, A., Xue, G., Lu, Z. L., Van der Linden, M., & Bechara, A. (2008). Neural correlates of envisioning emotional events in the near and far future. *Neuroimage*, *40*(1), 398-407.

De Brigard, F. & Gessell, B.S. (2016). Time is not of the essence: Understanding the neural correlates of mental time travel. In: Seeing the Future: Theoretical Perspectives on Future-Oriented Mental Time Travel. Oxford.

De Brigard, F, & Parikh, N. (2019). Episodic Counterfactual Thinking. Current Directions in *Psychological Science*. 28(1): 59-66.

De Brigard, F., Addis, D. R., Ford, J. H., Schacter, D. L., & Giovanello, K. S. (2013). Remembering what could have happened: Neural correlates of episodic counterfactual thinking. *Neuropsychologia*, *51*(12), 2401-2414.

De Brigard, F., & Giovanello, K. S. (2012). Influence of outcome valence in the subjective experience of episodic past, future, and counterfactual thinking. *Consciousness and Cognition*, *21*(3), 1085-1096.

De Brigard, F., Spreng, R. N., Mitchell, J. P., & Schacter, D. L. (2015). Neural activity associated with self, other, and object-based counterfactual thinking. *Neuroimage*, *109*, 12-26.

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

De Brigard, F. (2014). Is memory for remembering? Recollection as a form of episodic hypothetical thinking. *Synthese*. 191(2): 155-185.

De Brigard, F. (2019). Know-how, intellectualism, and memory systems. *Philosophical Psychology*. 32(5): 720-759.

De Brigard, F. (2020). The explanatory indispensability of memory traces. *The Harvard Review of Philosophy*. 27.

De Brigard, F. (2023). *Memory and Remembering*. Cambridge University Press.

Feldman, R. (1985). Reliability and justification. *the Monist*, *68*(2), 159-174.

Foster, D. J., & Wilson, M. A. (2007). Hippocampal theta sequences. *Hippocampus*, *17*(11), 1093-1099.

Furlong, E. J. (1948). Memory. *Mind*, 57, 16–44.

Gaesser, B., Sacchetti, D. C., Addis, D. R., & Schacter, D. L. (2011). Characterizing age-related changes in remembering the past and imagining the future. *Psychology and aging*, *26*(1), 80.

Gallistel, C.R. & King, A.P. (2010). Memory and the computational brain. Wiley-Blackwell.

Garry, M., Manning, C. G., Loftus, E. F., & Sherman, S. J. (1996). Imagination inflation: Imagining a childhood event inflates confidence that it occurred. *PBR*, 3, 208–214.

Gershman, S.J. (2023). The molecular memory code and synaptic plasticity: a synthesis. BioSystems, 224, 104825.

Gershman, S.J. (2023). The rational analysis of memory. In M. Kahana & A. Wagner (Eds.) Oxford Handbook of Human Memory. Oxford University Press.

Goldman, A., & Beddor, B. (2021). Reliabilist epistemology.

Hassabis, D., & Maguire, E. A. (2007). Deconstructing episodic memory with construction. *Trends in cognitive sciences*, *11*(7), 299-306.

Hassabis, D., Kumaran, D., Vann, S. D., & Maguire, E. A. (2007). Patients with hippocampal amnesia cannot imagine new experiences. *PNAS*, 104, 1726–1731.

Heil, J. (1978). Traces of things past. *Philosophy of Science*, 45(1), 60-72.

Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the national academy of sciences*, *79*(8), 2554-2558.

Huffman, D. J., & Stark, C. E. (2014). Multivariate pattern analysis of the human medial temporal lobe revealed representationally categorical cortex and representationally agnostic hippocampus. *Hippocampus*, *24*(11), 1394-1403.

Hume, D. (1748). *A treatise of human nature.* Oxford

Kleckner, I. R., Zhang, J., Touroutoglou, A., Chanes, L., Xia, C., Simmons, W. K., ... & Feldman Barrett, L. (2017). Evidence for a large-scale brain system supporting allostasis and interoception in humans. *Nature human behaviour*, *1*(5), 0069.

Kumaran, D., Hassabis, D., & McClelland, J. L. (2016). What learning systems do intelligent agents need? Complementary learning systems theory updated. *Trends in cognitive sciences*, *20*(7), 512-534.

Kurtzman, H. S. (1983). Modern conceptions of memory. *Philosophy and Phenomenological Research*, 44(1), 1–19.

Kwan, D., Craver, C. F., Green, L., Myerson, J., Gao, F., Black, S. E., & Rosenbaum, R. S. (2015). Cueing the personal future to reduce discounting in intertemporal choice: Is episodic prospection necessary?. *Hippocampus*, *25*(4), 432-443.

Lanzoni, L., Ravasio, D., Thompson, H., Vatansever, D., Margulies, D., Smallwood, J., & Jefferies, E. (2020). The role of default mode network in semantic cue integration. *NeuroImage*, *219*, 117019.

Leal Santos, S., Stackmann, M., Zamora, A. M., Mastrodonato, A., De Landri, A. V., Vaughan, N., ... & Denny, C. A. (2021). Propranolol decreases fear expression by modulating fear memory traces. *Biological psychiatry*, *89*(12), 1150-1161.

Liu, C., Yen, C. C. C., Szczupak, D., Ye, F. Q., Leopold, D. A., & Silva, A. C. (2019). Anatomical and functional investigation of the marmoset default mode network. *Nature communications*, *10*(1), 1975.

Loftus, E. F., & Pickrell, J. E. (1995). The formation of false memories. *Psychiatric Annals*, 25, 720–725.

Long, X., & Zhang, S. J. (2021). A novel somatosensory spatial navigation system outside the hippocampal formation. *Cell research*, *31*(6), 649-663.

Lyons, J. C. (2019). Algorithm and parameters: Solving the generality problem for reliabilism. *Philosophical Review*, *128*(4), 463-509.

Maguire, E. A., & Mullally, S. L. (2013). The hippocampus: a manifesto for change. *Journal of Experimental Psychology: General*, *142*(4), 1180.

Malcolm, N. (1963). *Knowledge and Certainty*. Cornell University Press.

Marr D. (1971). *Simple memory: a theory for archicortex. Phil. Trans. R. Soc. Lond. B 262, 23–81.*

Martin, C.B. & Deutscher, M. (1966). Remembering. *Philosophical Review*, 75: 161–196.

McAvan, A. S., Wank, A. A., Rapcsak, S. Z., Grilli, M. D., & Ekstrom, A. D. (2022). Largely intact memory for spatial locations during navigation in an individual with dense amnesia. Neuropsychologia, 170, 108225.

McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, 102, 419–457.

Michaelian, K., & Robins, S. K. (2018). Beyond the Causal Theory?: Fifty Years After Martin and Deutscher 1. In *New directions in the philosophy of memory*. Routledge.

Michaelian, K (2016). *Mental Time Travel: Episodic Memory and our knowledge of the personal past.* MIT Press.

Michaelian, K. (2022). Radicalizing simulationism: Remembering as imagining the (nonpersonal) past. *Philosophical Psychology*, 1-27.

Nader, K., & Einarsson, E. Ö. (2010). Memory reconsolidation: an update. *Annals of the New York Academy of Sciences*, *1191*(1), 27-41.

O'Keefe, J., & Recce, M. L. (1993). Phase relationship between hippocampal place units and the EEG theta rhythm. *Hippocampus*, *3*(3), 317-330.

Parikh, N., Ruzic, L., Stewart, G. W., Spreng, R. N., & De Brigard, F. (2018). What if? Neural activity underlying semantic and episodic counterfactual thinking. *NeuroImage*, *178*, 332-345.

Perrin, D. (2018). A case for procedural causality in episodic recollection. In New Directions in the Philosophy of Memory. Routledge.

Perrin, D. (2021). Embodied episodic memory: a new case for causalism?.

Race, E., Keane, M. M., & Verfaellie, M. (2011). Medial temporal lobe damage causes deficits in episodic memory and episodic future thinking not attributable to deficits in narrative construction. *Journal of Neuroscience*, *31*(28), 10262-10269.

Race, E., Keane, M. M., & Verfaellie, M. (2013). Living in the moment: Patients with MTL amnesia can richly describe the present despite deficits in past and future thought. *Cortex*, *49*(6), 1764-1766.

Raichle, M. E., MacLeod, A. M., Snyder, A. Z., Powers, W. J., Gusnard, D. A., & Shulman, G. L. (2001). A default mode of brain function. *Proceedings of the national academy of sciences*, *98*(2), 676-682.

Robins, S.K. (2016). Representing the Past: Memory Traces and the Causal Theory of Memory. *Philosophical Studies*, 173, 2993–3013.

Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words that were not presented in lists. *JEP:LMC*, 21, 803-814.

Romero, K., & Moscovitch, M. (2012). Episodic memory and event construction in aging and amnesia. *Journal of Memory and Language*, *67*(2), 270-284.

Rosenbaum, R.S., Cassidy, B.N., & Herdman, K.A. (2015). Patterns of preserved and impaired spatial memory in developmental amnesia. Frontiers in Human Neuroscience, 9, 196.

Rosenbaum, R.S., Priselac, S., Köhler, S., Black, S.E., Gao, F.Q., Nadel, L., & Moscovitch, M. (2000). Remote spatial memory in an amnesic person with extensive bilateral hippocampal lesions. Nature Neuroscience, 3(10), 1044–1048.

Rosenbaum, R.S., Stuss, D.T., Levine, B., & Tulving, E. (2007). Theory of mind is independent of episodic memory. Science, 318, 1257.

Rosenbaum, R.S., Gilboa, A., Levine, B., Winocur, G., & Moscovitch, M. (2009). Amnesia as an impairment of detail generation and binding: Evidence from personal, fictional, and semantic narratives in K.C. Neuropsychologia, 47, 2181–2187.

Rumelhart, D. E., McClelland, J. L., & PDP Research Group (Eds.). (1986). Parallel distributed processing: Explorations in the microstructure of cognition, Vol. 1: Foundations. MIT press.

Schacter, D. L., & Addis, D. R. (2007). The cognitive neuroscience of constructive memory: remembering the past and imagining the future. Philosophical Transactions of the Royal Society B: Biological Sciences, 362(1481), 773-786.

Schacter, D. L., Wagner, A. D., & Buckner, R. L. (2000). Memory systems of 1999.

Schacter, D. L., Benoit, R. G., De Brigard, F., & Szpunar, K. K. (2015). Episodic future thinking and episodic counterfactual thinking: Intersections between memory and decisions. Neurobiology of learning and memory, 117, 14-21.

Schroeder, M. P., Weiss, C., Procissi, D., Disterhoft, J. F., & Wang, L. (2016). Intrinsic connectivity of neural networks in the awake rabbit. Neuroimage, 129, 260-267.

Sforazzini, F., Schwarz, A. J., Galbusera, A., Bifone, A., & Gozzi, A. (2014). Distributed BOLD and CBV-weighted resting-state networks in the mouse brain. Neuroimage, 87, 403-415.

Skaggs, W. E., McNaughton, B. L., Wilson, M. A., & Barnes, C. A. (1996). Theta phase precession in hippocampal neuronal populations and the compression of temporal sequences. Hippocampus, 6(2), 149-172.

Spreng, R. N., & Andrews-Hanna, J. R. (2015). The default network and social cognition. Brain mapping: An encyclopedic reference, 1316, 165-169.

Szpunar, K. K., Watson, J. M., & McDermott, K. B. (2007). Neural substrates of envisioning the future. Proceedings of the National Academy of Sciences, 104(2), 642-647.

Talland, G.A. (1965). *Deranged memory: A psychonomic study of the amnesic syndrome*. Academic Press

Teyler TJ, DiScenna P. (1986). The hippocampal memory indexing theory. *Behav Neurosci* 100:147–152.

Tulving, E. (1985). Memory and consciousness. *Canadian Psychology*, 26, 1-12.

Van Hoeck, N., Ma, N., Ampe, L., Baetens, K., Vandekerckhove, M., & Van Overwalle, F. (2013). Counterfactual thinking: an fMRI study on changing the past for a better future. Social cognitive and affective neuroscience, 8(5), 556-564.

Vincent, J. L., Patel, G. H., Fox, M. D., Snyder, A. Z., Baker, J. T., Van Essen, D. C., ... & Raichle, M. E. (2007). Intrinsic functional architecture in the anaesthetized monkey brain. Nature, 447(7140), 83-86.

Vosgerau, G. (2010). Memory and content. *Consciousness and Cognition*, 19(3), 838-846.

Werning, M. (2020). Predicting the past from minimal traces: episodic memory and its distinction from imagination and preservation. *Review of philosophy and psychology*, 11(2), 301-333.

Wikenheiser, A. M., Gardner, M. P., Mueller, L. E., & Schoenbaum, G. (2021). Spatial representations in rat orbitofrontal cortex. Journal of Neuroscience, 41(32), 6933-6945.

Zhang, R., & Volkow, N. D. (2019). Brain default-mode network dysfunction in addiction. Neuroimage, 200, 313-331.