Accepted for publication in *Philosophical Psychology*. Please cite the published version.
https://doi.org/10.1080/09515089.2025.2571449

Memory traces and the (anti)causalism debate

José Carlos Camillo[ab]*

[a]*Faculdade de Filosofia, Universidade Federal de Goiás, Goiânia, Brazil;* [b]*Department of Philosophy II, Ruhr-Universität Bochum, Bochum, Germany.*

* Universidade Federal de Goiás, Programa de Pós-Graduação em Filosofia, Prédio de Humanidades II (Piso I), R. Jacarandá - Chácaras Califórnia, Campus Samambaia (Campus II), CEP 74690-900, Goiânia - Goiás - Brasil. E-mail: josecarloscamillo@gmail.com

Acknowledgements

I would like to express my sincere gratitude to Guilherme Ghisoni, Kristina Liefke, Jonathan Najenson, and James Openshaw for their invaluable comments and insights, which contributed significantly to the improvement of this work. I'm also deeply indebted to the editors of this special issue, Nikola Andonovski, Kourken Michaelian, and Sarah K. Robins, and to the anonymous reviewers who helped me enhance the quality of this paper.

Funding

This research was funded by FAPEG under the Grant Agreement No 202110267000414 and by DAAD under the funding program No 57710871.

Statements and Declarations

The author has no competing interests to declare that are relevant to the content of this article.

Accepted for publication in *Philosophical Psychology*. Please cite the published version.
https://doi.org/10.1080/09515089.2025.2571449

MEMORY TRACES AND THE (ANTI)CAUSALISM DEBATE

Abstract

One of the most important debates in the philosophy of memory is the causalism versus post-causalism debate. On the one hand, causalists defend the thesis that remembering requires a causal link via memory traces to a past experience. On the other hand, post-causalists deny that such causal link is necessary for remembering. Recently, some philosophers have suggested that Simulationism, particularly its anti-causalist thesis, is inconsistent with empirical findings about memory traces. In this paper, I analyze these findings and argue that although many of them favor Causalism and seem inconsistent with Simulationism, memory traces also provide data that are inconsistent with Causalism. For this reason, I defend that it is premature to dismiss any of these theories in light of the current evidence on memory traces.

Keywords: Episodic memory; memory traces; engrams; memory systems; Simulationism; Causalism.

1 Introduction

One of the most important debates in the philosophy of memory is the causalism versus post-causalism debate. On the one hand, causalists defend the thesis that remembering requires an appropriate causal link to a past experience. On the other hand, post-causalists deny that an appropriate causal link is necessary for remembering. Recently, some philosophers have suggested that the simulation theory of memory (STM), a post-causalist view, is inconsistent with empirical findings about memory traces (e.g. Robins, 2023, forthcoming, De Brigard, forthcoming, Andonovski, 2024).[1] As a consequence, they propose to reject STM and to return to the Causal Theory of Memory (CTM) (e.g. Robins, forthcoming) or at least a modification of the STM (De Brigard, forthcoming).

In this paper, I analyze the above-mentioned evidence and defend that, although many findings favor CTM and seem inconsistent with STM, there are some phenomena related to memory traces that CTM cannot account for perfectly, and that STM can account for. Given this dispute over the evidence, I will argue that it is premature to accept or reject either theory. Crucially, given the importance of the functional nature of memory

[1] Some philosophers make a distinction between engrams and memory traces (see Robins, forthcoming). In this paper, however, I will treat them as synonymous aligned with the scientific tradition (e.g. Kunz *et al*, 2018, Heinen *et al*, 2024).

traces for this theory choice, we need to gather more information about their functional nature in order to make a more informed choice.

The paper is organized as follows. In Section 2, I begin by answering a metatheoretical question in order to set out how I understand the causalism debate and what is at stake. Although many take it to be obvious and clear, the debate is not as well settled as some might think. I adopt the position that CTM and STM are hypotheses about the functional profile of the episodic memory system. Based on this metatheoretical discussion, I refine the CTM and STM hypotheses about the functioning of the episodic memory system (Section 3). Essentially, for CTM, the functional properties of the episodic memory system depend on the functional properties of the memory traces it manipulates. In contrast, STM rejects the functional characterization of memory traces. With these hypotheses clarified, I discuss evidence from memory traces and show how some findings favor CTM and are inconsistent with STM (Section 4). In Section 5, I turn to other memory trace findings and show that they are more consistent with STM than with CTM. I conclude in Section 6 with a cautionary note about rejecting either STM or CTM in light of recent evidence on memory traces.

2 What is the point of dispute in the causalism debate?

As stated at the outset of the paper, the CTM hypothesis posits the necessity of an appropriate causal link for remembering, a position that is rejected by the STM hypothesis. STM has two core theses (Michaelian, 2016, 2022, 2024, Michaelian et al., 2025):

Continuist thesis: remembering is a form of imagining because it results from the same underlying cognitive system that underlies imagining.

Anti-causalist thesis: a causal link to the past is not necessary for remembering.

At first sight, the disagreement between CTM and STM appears to be unambiguous, particularly because of the anti-causalist thesis. Nevertheless, the issue is not as straightforward as it may appear. The objective of classical CTM is to analytically define the concept of remembering (Martin & Deutscher, 1966). To provide such definition, Martin and Deutscher (1966) and Robins (2017) even considered possible creatures whose memory system was external to their body as having an evidential role

in their respective theories. In contrast, STM aims to provide a fruitful hypothesis for scientific contexts (Michaelian, 2016; Andonovski & Michaelian, 2024). Consequently, rather than wondering about imaginary creatures, STM defines remembering in terms of how human memory is likely to function, given current scientific evidence. As the aim and scope of CTM and STM are distinct, it is unclear why both cannot be successful simultaneously in their respective proposals.

In fact, recent papers have indicated that, due to these discrepancies in methodology, CTM and STM are not mutually exclusive. For instance, Craver (2020) put forth the idea that CTM is effective in defining remembering for epistemic purposes, while STM is effective in defining remembering for empirical purposes. McCarroll and colleagues (2024) recognize that CTM defines remembering in ordinary contexts, whereas STM defines it in scientific contexts. Openshaw (2023) interprets the debate as a non-necessary tension, proposing that STM is a hypothesis about the psychofunctional features of episodic memory, while CTM is a hypothesis of how remembering refers to particular past events. Finally, Schirmer dos Santos and colleagues (2023) view STM as a descriptive definition of remembering, whereas CTM offers a normative definition. In none of these interpretations of the debate there is a definitive point of contention between CTM and STM.[2] In particular, in the majority of these readings, CTM and STM are not in direct competition.

These interpretations of the debate are essential for preventing mere verbal disputes in philosophy of memory. It is important to note, however, that some recent versions of CTM are hypotheses that provide empirical characterizations of remembering in scientific contexts. They offer a psychofunctional theory of episodic memory and are descriptive rather than normative (e.g., Werning, 2020; Robins, 2023; De Brigard, forthcoming; Andonovski, 2024). If the aforementioned interpretations of the debate are accurate, it is reasonable to conclude that certain versions of CTM can be in conflict with STM. To uncover the tension, it is necessary to discover the specific points of contention and determine their relevance.

In terms of the relevance of the dispute, simulationists have recently asserted that STM is compatible with every instance of remembering being appropriately caused by the remembered experience (Andonovski & Michaelian, 2024). In another moment, Michaelian (2022) defends that one cannot find empirical evidence for necessity (see also

[2] To be fair, some of them have only Martin and Deutscher's version of CTM in mind when they defend that there is no direct tension between CTM and STM (e.g. Openshaw, 2023).

Michaelian et al., 2025). If the anti-causalist thesis is merely a modal thesis about how human memory could be, it is not pertinent to its objective of presenting a fruitful psychofunctional hypothesis for scientific contexts. Moreover, it is unclear how such a vague thesis could be empirically tested.

In light of recent metatheoretical discussions and in search of a relevant dispute, it is of the utmost importance to define the point of dispute. In this section, I will briefly consider three desiderata for defining this point. Firstly, both theses (causalist and anti-causalist) should have *clear empirical consequences*, aiming to avoid a loose modal interpretation of the debate (Desideratum 1). Secondly, both theses must be *competing*; that is, accepting one should lead to the rejection of the other (Desideratum 2). Ultimately, the point of contention must be *derived from current causalist/anti-causalist arguments* (Desideratum 3). This final desideratum simply requires that the point of dispute be derived from the existing debate, rather than introducing a new topic.

These desiderata align with Michaelian and colleagues' (2025, p. 11-12, emphasis added) observation that the dispute between CTM and STM requires clarification in naturalistic terms—specifically, by having clear empirical consequences (Desideratum 1). They state "a question arises about how he ought to conceive of the simulationism-causalism debate. If one can be a simulationist or a causalist without taking a stand with respect to NOCAUSE, it is far from *clear what distinguishes* the two families of theories." The italicized portion suggests that a proper characterization must present the theories as distinguishable and, in this sense, competing (Desideratum 2). Finally, Desideratum 3 aims to prevent a shift in the debate's focus, thereby avoiding the creation of an entirely new dispute.

One possible interpretation of the point of dispute is that CTM and STM are in disagreement regarding the appropriate characterization of mnemic phenomena. Characterizations are particularly useful in scientific contexts because they identify stable patterns in a phenomenon and the conditions under which such patterns occur (Bogen & Woodward, 1988; Colaço, 2018). Characterizations have clear empirical consequences, as they point to patterns that encompass the majority, if not all, instances of a phenomenon (Colaço, 2018). In this sense, the causalist thesis would be a strong thesis about virtually all instances of mnemic phenomena, thus satisfying Desideratum 1. In other words, the causalist thesis posits that virtually all mnemic phenomena exhibit appropriate causation, whereas the anti-causalist theory suggests that not all such phenomena display this type

of causal link.[3] On several occasions, it appears that STM is attempting to provide a characterization of mnemic phenomena. For instance, Michaelian (2016, p. 3) states that his objective is "an attempt to give an informative general characterization of remembering," and he views remembering as "a multifarious phenomenon" (2016, p. 5).

Nevertheless, characterizations are mainly delimitation of the explanandum and do not represent the explanans (Craver & Darden, 2013; Colaço, 2018, 2020; Siegel & Craver, 2024). In contrast, some recent versions of CTM aim at providing a more substantial explanation of remembering by appealing to the functioning of an underlying system or mechanism (e.g., Werning, 2020; Robins, 2023, forthcoming; De Brigard, forthcoming). To provide a concrete example, Werning (2020; see also Fayyaz et al., forthcoming) develops a functional model of the episodic memory system, which he equates with his version of CTM in opposition to STM. Similarly, the arguments for CTM presented by De Brigard (forthcoming), Robins (forthcoming), and Andonovski (2024) are also grounded in the functional characterization of the episodic memory system or its underlying mechanisms. In alignment with this objective, STM also invokes the operation of an underlying system (Michaelian, 2016, 2024). [4] For instance, Michaelian's primary argument against CTM relies on an analysis of the episodic memory system's function (cf. Andonovski & Michaelian, 2024), and he further classifies types of remembering based on how that system operates (Michaelian, 2021). Therefore, an interpretation framing the debate as a dispute over the correct characterization of remembering as a scientific phenomenon fails to satisfy Desideratum 3. In contrast, reading it as a debate about the functional properties of the episodic memory system successfully meets this requirement, as it is grounded directly in the arguments advanced by causalists and anti-causalists themselves. Similarly, Robins (2022) argues that, given the central role of the episodic memory system in current debates within the philosophy of memory, the focus has shifted. Rather than characterizing “remembering” as a superficial phenomenon, philosophy of memory “is now focused on characterizing *a neurocognitive system*, of which remembering is but a minor (and somewhat anomalous) component” (Robins, 2022, p. 5, emphasis added). As I just argued, this shift also encompasses the (anti)causalist debate.

[3] In this case, the anti-causalist thesis would not be true in a world where every instance of remembering is appropriately caused by the remembered experience.
[4] If Openshaw (2023) is correct in characterizing STM as a psychofunctional hypothesis, then the interpretation I follow is consistent with the objective of STM.

In light of the above, I concur with Andonovski (2021) in viewing the CTM versus STM debate as a dispute over the functional profile of the episodic memory system. A hypothesis regarding the functioning of a cognitive system may be weakly informative, with its empirical consequences being unclear, or it may be strongly informative, with clear empirical consequences. A weakly informative hypothesis can be described as a statement indicating that a system tends to function in a particular manner, with specific inputs and produces a predictable output. However, as Andonovski (2021, p. 143) states: "A theory which tells us that 'memories are caused by past experiences except when they aren't' would ring a lot of alarm bells and be rightfully treated as explanatorily suspect." Accordingly, if the CTM versus STM debate pertains to the functioning of the episodic memory system and aims at being strongly informative, these two theories should be regarded as hypotheses of exceptionless descriptions of a cognitive function.[5] This is due to the fact that, rather than focusing on the act of remembering, which is the result of intricate interactions between numerous cognitive systems, the emphasis is placed on a specific system, namely episodic memory. This approach produces an idealized isolation of the episodic memory system (Allott et al., 2021; Andonovski, 2023). That is, in contrast to presenting a hypothesis about the characterization of remembering, a superficial phenomenon, recent versions of the CTM versus STM debate narrow their focus to proposing a hypothesis about the functional profile of the episodic memory system. In doing so, it idealizes and simplifies the object of research (Andonovski, 2023). Consequently, a hypothesis about the functioning of a system is typically based on the assumption that it is a constant and inherent characteristic of the system.[6]

Such a strongly informative thesis about the functioning of the episodic memory system is crucial for satisfying Desideratum 1, that of having clear empirical consequences. In this way we can understand the scope of the modal thesis that anti-causalists deny. When Michaelian (2022) says that one cannot find empirical evidence for necessity, he probably has in mind an all-possible-worlds scope of necessity. However,

[5] De Brigard (2024) refers to this as "biological necessity," which he distinguishes from modal/metaphysical necessity.

[6] This represents a realistic interpretation of hypotheses pertaining to cognitive systems. In this view, cognitive systems are entities that are clearly delineated in both theory and reality. Nevertheless, many scholars challenge this interpretation and propose a more pragmatic understanding of cognitive systems. They argue that cognitive systems are not concrete entities in reality but rather useful fictions that help us predict and control certain aspects of cognition (see Chirimuuta, 2024). In this paper, I adopt a realist perspective on cognitive systems. However, a pragmatic approach would dismiss a hypothesis if it proves unable to facilitate the prediction and control of relevant phenomena. In this regard, CTM and STM can be effective for different purposes in the context of mnemic phenomena (see also Potochnik, 2017).

this is as empirically irrelevant as the thesis that water could be XYZ instead of H_2O and should not be an important part of a naturalistic account (Maddy, 2022). In other words, Michaelian sometimes considers the causalist thesis in the broadest modal sense, and consequently this thesis becomes empirically irrelevant. As a result, his own anti-causalist thesis also becomes empirically irrelevant because it is the negation of an empirically irrelevant thesis. Narrowing the modal scope of the causalist thesis has the effect of making both the causalist and anti-causalist theses empirically relevant. In this sense, understanding the thesis in terms of functional profiles allows us to limit the scope of this modal thesis in the (anti)causalist debate. In this case, the necessity claim refers only to the possible worlds in which the episodic memory system has the same functional profile as the actual world. As a consequence, the modal aspect of both the causalist and the anti-causalist theses becomes empirically relevant and can be tested in the terms I will discuss in the next sections.

If a hypothesis regarding the functional profile of a cognitive system is accurate, then, any instance of the system's functioning that does not align with the hypothesis (abnormal cases) would be considered as evidence inconsistent with the hypothesis (Pietroski & Rey, 1995). Abnormal cases here should not be taken in the traditional sense, but should refer within a theory to any instance of the functioning of the system that does not follow what is predicted by the theory (see De Brigard's example below). Such an understanding of systems is highly informative, as it allows for clear empirical consequences. That is, if those who support a particular hypothesis of a system's functional profile are unable to rule out inconsistent evidence, then the hypothesis is, at least, partially incorrect. By independently ruling out instances of inconsistent evidence, Pietroski and Rey (1995) mean that the reason for ruling out them is not due to the hypothesis itself, but rather some external evidence.

A recent example of rejecting a hypothesis in the face of unresolved inconsistent evidence comes from De Brigard's (2024) critique of Tulving's (2002) hypothesis regarding the profile of the episodic memory system. Tulving put forth the hypothesis that an autonoetic phenomenology is part of the functional profile of the episodic memory system (see Andonovski, 2023). De Brigard, however, presents an argument against this proposal, noting that there are numerous instances of episodic memory functioning that do not include autonoetic phenomenology. These cases, he asserts, cannot be dismissed as the result of external interference. Consequently, De Brigard concludes that autonoesis should be excluded from the functional profile of the episodic memory system. In other

words, abnormal cases like these are seen as inconsistent evidence for a hypothesis. In theory/hypothesis choice, one should favor theories that can explain away inconsistencies (Kitcher, 1990, Parkkinen et al., 2017). Since Tulving's hypothesis could not rule out these empirical inconsistencies as resultant from external interferences, De Brigard argued that we should reject this hypothesis.

One problem with this theory choice strategy is that the inability to rule out inconsistent evidence may not be a problem with the hypothesis/theory being rejected, but may be due to shortcomings in the methodological and technological approaches employed. Indeed, theories that were eventually accepted by the scientific community faced a long history of inconsistent evidence before it could be ruled out (Kitcher, 1990). For example, continental drift faced the theoretical argument that there was not enough force in the Earth's magma to move continents (Santana, 2021); the cholesterol hypothesis of heart disease faced a long period of clinical trials in which interventions to control cholesterol did not affect the likelihood of death from heart disease (Parkkinen et al., 2017); the bacterial hypothesis of peptic ulcer disease faced the problem that no bacteria were found in the human stomach even in a sample of more than a thousand biopsies (Zollman, 2010). Over time, however, all of these hypotheses have gained unanimous acceptance within the scientific community.

Given that many of today's accepted hypotheses faced evidence for decades that was considered inconsistent with them, we should be cautious about rejecting a hypothesis/theory solely in light of inconsistent evidence (Kitcher, 1990, Zollman, 2010, De Cruz & De Smedt, 2013). For this reason, it is common for competing theories to persist for long periods of time, even in the face of conflicting evidence (Kitcher, 1990, Pietroski & Rey, 1995, Zollman, 2010). I will return to this point in the conclusion, where I will argue that, in light of the evidence discussed throughout this paper, both CTM and STM have inconsistent evidence regarding memory traces, and that we should not be too quick to dismiss either of them.

For now, understanding the (anti)causalism debate as a debate over the correct functional profile of the episodic memory system is arguably the right way to interpret the tension between at least some versions of CTM and STM. Firstly, this understanding allows for clear empirical consequences by presenting what can be seen as inconsistent evidence or evidence in favor of one theory (Desideratum 1). Moreover, it would also make CTM and STM competing theories (Desideratum 2). The former claims that appropriate causation is a crucial part of the functional profile of the episodic memory

system, whereas the latter claims that appropriate causation is not a component of the functional profile of the episodic memory system (see next section). Furthermore, it is consistent with the STM's emphasis on the episodic system and with recent CTM versions that seek to delineate the functional profile of the episodic memory system (as discussed in this section) (Desideratum 3). Therefore, the CTM versus STM debate, which concerns the correct hypothesis regarding the functional profile of the episodic memory system, satisfies the three desiderata for a point of dispute previously outlined in this section. Hence, I will adopt this reading of the debate for the remainder of the paper.

In this section I revisited the causalism debate in philosophy of memory. I clarified what is the point of dispute between some versions of CTM and STM, namely, the functional profile of the episodic memory system. CTM and STM provide competing hypotheses of this functional profile. As such, they should be able to rule out inconsistent evidence for their hypothesis, particularly data that are well explained by the competing theory. If we accept that this is the point of dispute, we should clearly define the hypothesis that CTM and STM are making regarding the functioning of this system and how exactly they oppose one another, which I will do in the next section.

3 Clarifying CTM and STM as hypotheses about the functional profile

Although some propose a naturalistic interpretation for the classical CTM (De Brigard, 2020), it is noteworthy that the goal of Martin and Deutscher (1966) was different from hypothesizing the functional profile of the episodic memory system.[7] However, as seen in the previous section, some recent versions of CTM aim to provide such a hypothesis.

Before proceeding, it is useful to clarify what counts as evidence for the functional profile of a cognitive system in general, and of episodic memory in particular. Establishing such a profile is a difficult task that requires multiple types of evidence (cf. Mahon & Cantlon, 2011; Friston & Price, 2011), and in the case of episodic memory, we are likely still far from achieving it. A common approach is to first demonstrate that a system is functionally dissociable or independent from other systems, and then examine how its internal operations—whether computational, algorithmic, or implementational—contribute to its function (Sternberg, 2011; Coltheart, 2011). In other words, once an (at

[7] For instance, episodic memory system was only hypothesized six years later (Tulving, 1972). Moreover, they were concerned with a theory of remembering, phenomenal features, not of its underlying structures—for the difference, see Allot *et al* (2021) and Siegel and Craver (2024).

least partially) independent function has been identified, a functional analysis can be conducted to determine how the system performs that function.[8]

An absolute functional profile, however, is beyond the current capacity of cognitive (neuro)science (cf. Shallice, 1988; Sternberg, 2011; Mahon & Cantlon, 2011; Coltheart, 2011, 2017). What can be established with greater confidence is a comparative profile, based on comparative evidence (Sternberg, 2011). This involves investigating whether two systems are functionally independent by examining differences in task or behavioral performance between them (Tulving, 1984; Sternberg, 2011), as well as differences in the operations underlying these performances (Bergeron, 2016). Such evidence provides not a complete profile but a partial, comparative one.[9] In other words, we can dissociate the operations of one system from those of another, though we cannot fully characterize either system in absolute terms, that is, without comparing it with another putative system. Even with this limitation—the inability to produce a complete functional profile—we can develop a robust approach. This method would combine evidence of dissociations in task performance and neurobiological implementation with a functional analysis of the underlying mechanisms and processes (Bergeron, 2007, 2016).

Tulving's work illustrates this methodology well (cf. 1983, 1984, 1986, 1999; Schacter & Tulving, 1994). He argued for the dissociation of semantic and episodic memory systems on the basis of both task performance (1983, 1984) and neurobiological evidence, such as the case of K.C. (Rosenbaum et al., 2005). Beyond these convergent dissociations, he also pursued a functional decomposition of memory systems, considering not only their neural substrates but also their operations and operational rules (Schacter & Tulving, 1994; Tulving, 2002). Still, he acknowledged that a complete account of the functional profile of memory systems was far from achievable (e.g., 1984, 1999). Even today, decades after Tulving's retirement, the functional characterization of episodic memory remains unsettled (cf. Andonovski, 2023; De Brigard, 2024).

[8] This sequence is merely logical. In practice, scientists use functional analysis to identify functions, and the reverse is also true: discovering a dissociable function contributes to the process of functional analysis (Bergeron, 2007).

[9] Interestingly, as we will see, Tulving (1983, 1984, 1986, 1999) consistently framed this issue in relation to memory systems. He argued that a precise characterization of the number and functions of these systems is beyond our current understanding. In his view, research can identify dissociations between behavioral tasks and neurobiological differences, demonstrating that certain memory systems are distinct. However, he maintained that a full description of their specific functions and operational mechanisms would require far more evidence than was available in his time—and likely remains so today.

Despite these limitations, Tulving maintained that existing evidence was sufficient to distinguish episodic memory from other systems, such as semantic and procedural memory (1983, 1984, 1999). For him, a key point of dissociation—and thus a crucial marker of episodic memory's functional profile—was the functional features of the engrams or memory traces used by the different systems. Addressing the core question of how memory systems can be individuated, he remarked: "The controversy materializes when we pose the next apparently innocent question: where is the information about the […] event stored?" (1999, p. 18). In this sense, for him, the hypothesis that there are multiple memory systems

> means that there is no 'single' engram, or single memory traces, or single representation of the event […] As a result, different kinds of information, representing the many different aspects of the event, are stored at different independent 'storage sites.' These sites correspond to memory systems and their subsystems (1999, p. 20).

Thus, while the type of engrams or memory traces a system operates over is not its only crucial feature, Tulving considered it central evidence for functionally characterizing memory systems. We can think of many reasons for the importance of engrams, or memory traces. According to Schacter & Tulving (1994), memory systems should be characterized by three criteria: a distinct class of operations, a specific set of properties (including neural substrates, type of information, and rules of operation), and convergent dissociations. Engrams are directly relevant to the first two criteria, as they are distinguished by the type of information they represent and they possess distinguishable neural substrates. Consequently, investigating engrams could provide crucial information about the operations performed on these traces and the rules that govern those operations.

Given the interpretation of the causalism–anticausalism debate adopted in this paper—namely, that the dispute concerns the correct functional profile of episodic memory—it is unsurprising that evidence about memory traces plays a central role in the arguments. Indeed, not only is STM concerned with the nature of memory traces from its inception (De Brigard, 2014; Michaelian, 2016; cf. Michaelian et al., 2025), but traces also feature prominently in recent causalist accounts, as will be evident in the following

discussion (cf. Robins, 2016, 2023, forthcoming; Werning, 2020; Andonovski, 2024; see also De Brigard, 2020, forthcoming).[10]

According to these causalist views (e.g., Werning, 2020, Robins, 2023, forthcoming, De Brigard, forthcoming, Andonovski, 2024), the episodic memory system encodes, stores, and retrieves (i.e., manipulates) discrete representations of particular events. These representations are called memory traces and are formed at the moment of experiencing the event. There are some differences between these proposals. Werning (2020) and De Brigard (forthcoming), for example, do not consider memory traces to be representations, but merely *hippocampal indexes*. Still, these hippocampal indexes are formed during events (comparable to encoding), remain stable in the hippocampus over time (comparable to storage), and are reactivated to contribute to remembering (comparable to retrieval). Crucially, these indexes are also specific to each experienced event (or at least to those relevant enough to form such indexes). It is safe to conclude, then, that these versions of CTM functionally describe the episodic memory system in terms of the functional properties of the memory traces that constitute the representational vehicles that the system manipulates.

Importantly, they assume or defend that storage somehow involves maintaining the original identity of the trace.[11] By identity maintenance, I mean that the trace represents the same event over time, even with some possible differences in its content. Andonovski (2024), for example, argues that despite some changes in its content, traces maintain their identity as references to particular events that formed them. Robins (2016, 2017) also assumes that traces retain such identity, despite neural changes (Robins, 2020, 2023, forthcoming). De Brigard (2020, forthcoming) also states that traces maintain their particularity over time, which is arguably the same as what I am calling identity here.[12]

[10] This perspective also aligns with recent proposals emphasizing the critical role of storage processes in memory theory (Andonovski, 2021) and advocating for its incorporation into the debate between CTM and STM (Andonovski, 2022; Openshaw, 2023; Robins, forthcoming).

[11] It is important to clarify that a trace's identity is tied to its functional role of representing a specific event. Changes in its physical implementation (cf. Robins, 2020) or even its content (Andonovski, 2024) do not alter this core representational function. Therefore, from a functional perspective, the trace remains the same despite these underlying changes.

[12] Some might contend that "identity" is an imprecise term, suggesting a "coreference relation" is more accurate. However, I retain the term "identity" for two reasons: it is the standard term used in the scientific literature (e.g., Choucry et al., 2024), and it accurately reflects a functional understanding of memory traces. From this perspective, a trace maintains its identity by preserving its core functional role—storing information about a specific event—as long as it continues to refer to that event. Therefore, it remains the same trace despite changes in its physical implementation or even its specific content, as argued in the previous footnote.

For this reason, it makes sense to summarize this hypothesis as a two-sided hypothesis:

Causalist Hypothesis (CH):

- *CH_1)* the episodic memory system manipulates (encodes, stores, and retrieves) discrete representations (or indexes) of particular events, named memory traces;
- *CH_2)* the episodic memory system somehow secures the identity of these traces over time, i.e., they continue to represent the same event even under some modifications;
- *CH_3)* these traces are formed during the experience of the particular event they represent, a process known as encoding.

In sum, CH as a whole maintains the notion of appropriate causation from Martin and Deutscher's (1966) version of CTM because it posits a persistent causal link between the past experience and the recollection via memory traces. Moreover, such traces exhibit specificity/particularity/monogamy (Robins, 2016, Sant'Anna, 2021, Langland-Hassan, 2022) and they are contiguous/non-deviant (Bernecker, 2017, Robins, 2017). If we understand the above mentioned naturalistic versions of CTM as hypotheses about the functional profile of the episodic memory system (Section 2), then we should see CH as proposing a hypothesis according to which the episodic memory system forms traces (discrete representations) in response to the experience of particular events, and these traces are manipulated by that system to retrieve information about the particular event the trace represents. Crucially, if a particular retrieval involves the contribution of the episodic memory system, that memory experience will have appropriate causation, as stated by traditional CTM.

STM proposes a hypothesis that competes with CH_1-CH_3. As Andonovski and Michaelian (2024) put it, the core claim of STM is that the episodic memory system is the same cognitive system that produces episodic imagining (continuist thesis). Consequently, it cannot have the functional profile of CH, because such a functional profile would not allow imagining counterfactual events.[13] In particular, imagining

[13] De Brigard (2023, forthcoming) claims that these two hypotheses (CH and continuism) are compatible, but he does not explicitly develop an account in which this integration works. That is, there is no clear account of the explanatory role that memory traces as defined by CH (specific/particular memory traces) play in episodic construction. Michaelian (2016) also thinks that they are not inconsistent with the continuist thesis. He defends that these specific/particular memory traces sometimes participate in episodic simulation

counterfactual events does not depend on a causal trace formed during the experience of those events (Michaelian, 2016, 2022, 2024). In favor of STM, simulationists appeal to abnormal cases for CH_1-CH_3.[14] For example, some argue that memory consolidation involves generalizations (Moscovitch & Gilboa, 2024, Aronowitz, forthcoming) that lead to a loss of specificity/particularity of representations, thus contradicting CH_2. Moreover, episodic memory is involved in the recall of unexperienced, "false" events (Bernstein & Loftus, 2009), apparently contradicting CH_3. Finally, rejecting CH would better explain why memory is so frequently and easily distorted even in healthy patients (Schacter & Addis, 2007). The anti-causalist thesis of STM can be understood as mere rejection of CH. That is most of the time what Michaelian (2016, 2022, 2024) does (see footnote 13). However, other "simulationists", mostly from psychology have a more positive hypothesis to defend. In particular, Addis and Schacter (2007), Addis (2018, 2020), Hassabis and Maguire (2009), Barry and Maguire (2019a) (see also Ménager *et al*, 2021, 2022, Langland-Hassan, 2022), among others agree that

Simulationist Hypothesis (SH): the episodic system manipulates schemas of events (or representations of types of events rather than tokens) along with sparse semantic information about those events (e.g., about objects, actions involved in those events) to create simulations/scenes/scenarios of events.[15]

Thus, SH, rejects CH_1, the particularity aspect of CH. In other words, the type of representational vehicle that the episodic system manipulates is not of particular events, but about event types, such as schemas (Barry & Maguire, 2019a, Addis, 2020). Consequently, there is no particular identity to be maintained (contra CH_2) and no particular/specific formation of traces (CH_3). This hypothesis about the functional profile of the episodic system would better account for the evidence mentioned above that are inconsistent with CH. In particular, the use of schemas and other concrete details would be crucial for imagination, would involve generalizations, would explain why we

construction and sometimes do not. However, as seen in the previous section, without specifying how this works, such a claim is explanatorily suspect (Andonovski, 2021, p. 143).

[14] One can notice the similarity of this argument and De Brigard's argument in the previous section as an exemplar of appeal to inconsistent evidence as a reason to reject a hypothesis.

[15] Some of the psychologists who defend this hypothesis call this system the [scene] construction system or the simulation construction system.

remember unexperienced events, and why memory is so often distorted (see Schater et al., 2011, 2012).[16]

In the next section, I present some findings from research on memory traces that threaten SH and favor CH.

4 Memory traces, evidence inconsistent with STM and the comeback of CTM

Evidence for CH comes from studies of memory traces (or engrams). Memory traces are understood as permanent synaptic or molecular modifications that store information about specific events and are formed during the experience of these events (Asok et al., 2019, Josselyn & Tonegawa, 2020). In this section, I will briefly summarize three lines of research that support CH: computational modeling, neuroimaging studies, and molecular research.[17]

Some studies have built computational models of both CH and SH. In running the models, researchers found that SH performed poorly compared to human performance, whereas models that followed CH matched human performance (Ménager et al., 2021, 2022). Importantly, despite SH's focus on memory distortions, human memory has generally been measured to be about 70% accurate under laboratory conditions (Powers et al, 1979, Loftus et al, 1992) and is even more accurate under naturalistic/ecological conditions (Diamond et al, 2020, Foudil et al, 2021, Finley & Brewer, 2024). Models following SH have failed to replicate such levels of accuracy in computational tests.

The second type of studies are neuroimaging approaches to memory traces (e.g., fMRI, EEG, MEG). These neuroimaging studies have found discrete representations for specific events in the hippocampus (Kunz et al, 2018, Xue, 2022, Ranganath, 2024, Heinen et al, 2024). Crucially, the reactivation of these discrete representations is associated with the recall of these specific events. These neuroimaging techniques analyze patterns of neural activity and cannot precisely localize traces (see Kunz et al, 2018, Roskies, 2021). However, they can serve as epistemic proxies for the existence of these discrete traces (Roskies, 2021, see also Klein, 2010), i.e., because different events have different patterns of neural activation, it is likely that traces (molecular or synaptic modifications—Gershman, 2023, Colaço & Najenson, 2023) underlie these patterns and

[16] As seen in the characterization of SH, it ultimately rejects the existence of an episodic memory system because, if it were true, it would not be a cognitive system dedicated to memory, but to scene construction or simulation (Andonovski et al., 2024).
[17] Andonovski and colleagues (2024) have provided a similar summary of research that, in their view, is inconsistent with SH and supportive of CH.

store information about these specific events.[18] However, this is an educated guess.[19] It is also possible that there is no storage (as proposed by Werning, 2020, and De Brigard, forthcoming).

More important for CH is the prediction that if the episodic memory system were to manipulate discrete representations of particular events, it would show pattern separation (McClelland et al., 1995). Namely, similar events (e.g., similar spatiotemporal location) will show similar patterns of neural activity, while dissimilar events will show dissimilar patterns. This leads to the concern that the system may not be able to distinguish between the representations of very similar events. Consequently, very similar events would be treated as the same by the system, losing the identity hypothesized by CH_2. Pattern separation was the prediction that in order to ensure trace specificity and avoid interference between events that are similar above a threshold, the hippocampus would produce pattern separation, whereby very similar events would be represented by very dissimilar patterns. In sum, the greater the similarity of events, the greater the similarity of patterns, unless this similarity exceeds a threshold (very similar), in which case the patterns would be very *dissimilar* to avoid interference and loss of identity. Pattern separation has been widely reported using different neuroimaging techniques (see Yonelinas et al, 2019, Clewett et al, 2019, Morales et al, 2020, Brunec et al, 2020, Theves et al, 2024). This phenomenon is evidence that the episodic memory system is indeed concerned with maintaining the identity of representations of specific events (as seen in CH_2).

Finally, the third type of study that has provided evidence for CH comes from biomolecular research. Using techniques such as transgenics, optogenetics, and chemogenetics (Kunz et al., 2018, Josselyn & Tonegawa, 2020), scientists have found discrete memory traces that represent specific events. Below, I provide a brief summary of these findings and how they support CH. [20]

- *Trace identification*: when one experiences an event, specific neurons representing that event in the neocortex and hippocampus become more excited due to stronger activation (Josselyn & Frankland, 2018). This results in the

[18] Patterns of neural activity are not directly indicative of these modifications. However, the stability of the patterns suggests that stable modifications have been implemented in specific neurons.
[19] Roskies (2021) calls these patterns epistemic proxies for neural representations for the reason given in footnote 18.
[20] For more substantial overviews, see Najenson (2021), Robins (2023, forth.), and Andonovski (2024).

production of immediate early genes (a set of genes formed after such strong activation) that are used to track these specific neurons (i.e., serve as tracing markers) (Tonegawa et al., 2018, Choucry et al., 2024). The neural assemblies tracked in this process are referred to as traces or engrams.

- *Trace manipulation*: these traces are then manipulated to better understand their causal profile using two different strategies: loss-of-function and gain-of-function (Kunz et al, 2018, Josselyn & Tonegawa, 2020). The loss-of-function strategy uses light, chemical, or genetic manipulation to block the activation of these particular neurons. Gain-of-function activates these assemblies using similar techniques.
- *Results*: Of relevance to this work is the finding that these traces store information about specific events. Crucially, they are necessary for remembering information about these events - loss-of-function manipulations result in an inability to remember only the particular event associated with the trace. Moreover, they are also sufficient for remembering information about these events—gain-of-function manipulations result in remembering the particular event associated with the trace, regardless of contextual cues. Moreover, these neural assemblies are formed during the experience of a specific event and exhibit molecular transformations that allow them to retain information about the event even as systematic brain transformations occur (Tonegawa et al., 2018, Asok et al., 2019, Josselyn & Tonegawa, 2020), suggesting a type of storage and maintenance of identity (Nambu et al., 2022, Gershman, 2023).
- *Consequences*: Consider that neural representations are patterns of neural activity that reliably covary with a specific stimulus, that have downstream effects directed toward that specific stimulus, and that can be activated in the absence of the stimulus, thereby maintaining its downstream effects (Vilarroya, 2017, Baker et al, 2022, Cao, 2022, Piccinini, 2022). In this case, traces can be understood as discrete representations of specific events because they reliably covary with these events, have downstream effects in respect to these events, and can be activated in the absence of the event while maintaining the same downstream effects. If one understands the molecular transformations that these traces undergo as storage processes, then one can confidently conclude that memory traces are discrete representations of particular events, and that they are used to encode, store (maintain their identity), and retrieve information about these particular events, just as described by CH (cf. Robins, 2023, forthcoming, Andonovski, 2024).

The three lines of research reviewed in this section provide strong evidence in favor of CH. They constitute robust evidence that the episodic memory system manipulates discrete representations of particular/specific events (CH_1). This particularity persists over time, maintaining identity (CH_2) and these discrete representations are formed during the experiences they represent (CH_3). However, as seen in Section 2, a hypothesis about the functional profile of a cognitive system should be able to independently rule out inconsistent evidence. For this reason, CH should also be able to rule out the the inconsistent evidence discussed at the end of Section 3. To some extent, causalists may succeed in explaining away these empirical inconsistences with their theory, such as memory distortions, generalization processes, memory for unexperienced events, and the relationship between memory and imagination.

Firstly, as seen in the beginning of this section, memory performance, including its distortions, is generally better predicted by CH models than by SH models. Secondly, generalized information about events is processed in different brain regions than those that process information about specific events. For example, posterior medial regions in the cortex are associated with discrete representations of particular events, whereas anterior temporal regions are associated with generalizations of events (Ranganath & Ritchey, 2012, Ritchey et al., 2015, Ritchey & Cooper, 2020, Ranganath, 2024). Although these regions show strong anatomical connections, they are not as well connected to each other as they are internally connected, suggesting that they are functionally individuated (Ritchey & Cooper, 2020). In addition, different hippocampal regions appear to process either generalized or specific information about events (not both) and use different markers of immediate early genes to compute general versus specific information about events (Choucry et al., 2024, Yelhekar et al., 2024). In light of all this evidence, CH can consider generalizations to be produced by systems dissociated from the episodic memory system and maintain that the episodic memory system is specialized for particular events. Finally, there are several dissociations between episodic memory and episodic imagination that are better explained by hypotheses that dissociate them than by hypotheses that integrate them into the same cognitive system (for a review, see Mullally & Maguire, 2014, Schacter et al., 2012, Camillo, 2025). [21]

[21] An abnormal case mentioned by Langland-Hassan (2022, p. 11), which I did not mention in section 3, but which can also be ruled out by CH, is the case he calls forgetting-by-repetition. This is the common phenomenon in which one fails to remember a particular event because one has experienced many events similar to the one she is trying to remember. Langland-Hassan argues that CH cannot invoke interference

As seen in this section, research on memory traces suggests that they have the functional profile predicted by CH. Moreover, CH can, to some extent, rule out inconsistent data by appealing to independent evidence. In other words, these data can be interpreted as the result not of the episodic system, but of other systems. This motivated philosophers to reject STM (or, at least, its anti-causalist thesis) (e.g. Robins, forthcoming, Andnovski, 2024, Andonovski et al., 2024, De Brigard, 2024). However, there are also other interesting findings about memory traces that remain inconsistent with CH and that may so far favor SH, as I will discuss in the next section.

5 Memory traces and evidence inconsistent with CTM

In this section I discuss some phenomena associated with memory traces that are inconsistent with CH. CH has not yet ruled them out and, crucially, they can be seen as evidence for SH. I begin by discussing a case of remembering unexperienced events, as found in Najenson (forthcoming) (Section 5.1). Then, I revisit evidence for generalization in episodic memory in consolidation (Section 5.2) and co-allocation processes (Section 5.3).

5.1 Memory traces of unexperienced events

Recently, Najenson (forthcoming) discussed the consequences of some fascinating studies on memory trace implantation for CH_3. Briefly, Vetere and colleagues (2019) conducted a striking experiment focusing on two neural assemblies in mice. The first group consisted of sensory neurons that respond to the smell of a particular substance, acetophenone. The second group consisted of dopamine neurons in the medial ventral tegmental nucleus, which are known to play a key role in fear responses. Importantly, the mice used in the study had never previously encountered the odor of acetophenone. By activating both groups of neurons at the same time, the researchers were able to artificially link the perception of that smell to a fear reaction, forming an engram of fearing it. As a result, the mice displayed fear when they later encountered acetophenone in their environment, even though they had no prior experience with it. Replications of these results using different sensory neurons—from visual, auditory, and somatosensory

to explain this phenomenon because if the traces were discrete, there would be no interference. In fact, as we have seen, pattern separation and other properties of memory contribute to the maintenance of identity and the avoidance of interference. However, interference occurs not because the trace representations are poorly dissociated, but because the cues are ambiguous and insufficient to select a particular trace (Ekstrom & Yonelinas, 2020). For further criticism of Langland-Hassan's argument, see also Álvarez (2025).

cortices—were also successful (Luis-Islas et al., 2022). As Najenson (forthcoming) discusses, these cases provide evidence that memory traces can be formed without previous experience, which is inconsistent with CH_3 because it describes the formation of memory traces as causally dependent on the events they represent.

Causalists may object that this is the result of an external intervention (optogenetic intervention), that the system involved is not the episodic system but something like the reward or limbic system, that this trace is not of an event, among other objections. These objections may be problematic for causalists, however, because most biomolecular research with memory traces (Section 4) follows similar protocols. For example, the interventions used in this study were gain-of-function interventions that provided evidence for CH. The system involved here (reward/limbic) is involved in most, if not all, of the biomolecular studies mentioned in Section 4. Consequently, if these neural assemblies are not used by the episodic memory system, then it is quite likely that the assemblies that provided evidence for CH are not part of the episodic memory system either.

Moreover, as Najenson (forthcoming) argues, the functional and structural similarities between natural memory trace formation (see Josselyn & Tonegawa, 2020) and these cases of memory trace implantation are strong enough to convince us that these interventions are informative about how memory traces work in natural circumstances. Furthermore, Najenson provides evidence that engrams can form spontaneously in natural contexts, without being linked to a specific experience. This indicates that experimental interventions which create new, experience-independent traces are replicating a natural phenomenon rather than producing an artifact. Indeed, not only are memory traces formed naturally from activations similar to those that the optogenetic interventions were designed to simulate, but also the updating of memory traces suggests that reconsolidation processes are similar to this case of false memory induction (Gershman *et al.*, 2014). Moreover, the response behavior of artificial memory and natural memory is very similar (Park *et al.*, 2024). For this reason, we can reliably extrapolate memory trace implantation studies described above to the formation of episodic memory traces in humans (Steel, 2008).

Hence, causalists either reject a large group of studies that have been taken to favor them, or they explain why memory trace implantation is different from the other studies. This explanation, it is important to emphasize, must be based on external evidence and not on CH itself. Nonetheless, given the similarities between the processes

of memory trace implantation and how scientists investigate memory trace formation and its dynamics, one can consider it to be unlikely that CTM will succeed in explaining away the case of memory trace implantation.

5.2 Consolidation, generalizations and the transience of memory traces

Given the importance of memory traces' functional features to the (anti)causalist debate (see Section 3), investigating systems consolidation is crucial. This process can reveal how event information becomes more permanently stored in the brain over time (Asok et al., 2019). Such an investigation would provide a more clear and complete understanding of traces' functional features.

According to the scene construction consolidation theory, memory traces for specific events, such as those posited by CH_1, exist only for recent events. After a few weeks or months, the information is transferred to cortical areas and generalized, losing its identity as it no longer represents a particular event (Barry & Maguire, 2019b, Aronowitz, forthcoming). According to this theory, the original trace is erased or fades in the process, and because the information has been generalized, the identity is lost, which is inconsistent with CH_2. In other words, their theory posits that specific memory traces, like those hypothesized by CH, are highly transient. According to this view, retrieving a memory after a significant time lapse depends on reconstructing past events using schemas and generalizations—a process akin to inferring particular events from general schemas (Barry & Maguire, 2019b). This perspective aligns well with SH's hypothesis (cf. De Brigard, 2014; Michaelian, 2016).

In fact, if this theory is correct,[22] the functional profile of episodic memory could be described in one of the following ways: 1) manipulate different types of representations over time (representations of specific events for recent memories and representations of generalizations for older memories); 2) manipulate event representations regardless of the specificity/singularity of the event represented (Andonovski, 2020, 2022); 3) not be a proper system, i.e., be indistinguishable from semantic memory systems (Aronowitz, forthcoming); 4) be a system responsible for processing only recent memories of specific events (Vecchi & Gatti, 2020). All of these possibilities are inconsistent with CH_1 and CH_2 because there is a necessary loss of particularity and identity after a few weeks or

[22] I will present supporting evidence in what follows. However, it is important to note that many scientists consider a competing theory, the Trace Transformation Theory (Xue, 2022; Moscovitch & Gilboa, 2024), to be better supported by evidence. For a detailed debate on the evidence, see Squire et al.; Moscovitch and Nadel (2019) and Barry and Maguire (2019a).

months. Consequently, CH should at least be modified. More importantly, maintaining the identity of the trace over time would definitely not be part of the functional profile of episodic memory, precisely because the initial specific trace would be erased, thus contradicting CH_2.

In response, Andonovski (2020) suggests that CTM could give us an account of appropriate causation in generalized (non-singular) memories.[23] Indeed, he recently provided evidence that event individuation can lead to the integration of temporally contiguous events as the same event (Andonovski, 2024). In other words, the perceptual system and working memory individuate events in a variety of ways, and often longer events are individuated as particular, while at other times shorter events are individuated as particular. In this case, general events (longer events, such as an entire day or week) may be individuated as particular by cognitive systems. This would ensure CH_1, in which episodic memory would still only manipulate representations of particular events.

However, this would not solve the problem of the transience of the identity of traces over time, which is inconsistent with CH_2. Some who have responded to this problem have argued that the evidence presented in Section 4 strongly suggests that memory traces of particular events are not transient (e.g. Moscovitch & Nadel, 2019, Moscovitch & Gilboa, 2024). Nonetheless, Barry and Maguire (2019a, 2019b) note that the evidence for the persistence of the identity of memory traces over time comes from studies that only track traces for up to four weeks or thirty-one days. According to their hypothesis, specific memory traces fade up to four months after encoding, when these traces would no longer exist and recall would depend entirely on generalized information. In other words, after these four months, the specific memory trace disappears, and episodic memory would function exactly as described by SH. In this sense, old episodic memories are only constructed through the recontextualization (i.e., the creation of a new spatiotemporal context via hippocampal activity) of generalized (decontextualized) information (Barry & Maguire, 2019b). According to them, evidence against their view has not been provided by the studies discussed in Section 4 because these studies are limited to one-month engrams. Supporting their theory, research shows that pattern separation in the hippocampus—which ensures computational specificity for particular

[23] In fact, Andonovski (2020) argued that generalization processes suggest that the CTM versus STM debate should be reformulated, since in his view the episodic memory system does not exhibit singularism (what I call specificity/particularity in this paper). However, as seen in the previous, Andonovski's interpretation of the implications of generalizations is not the only possible hypothesis. Moreover, the integration of non-continuous events seems to contradict a reformulated version of CH (see the following discussion).

memories (see Section 4)—vanishes for memories aged six to twelve months (Theves et al., 2024). This finding strongly suggests that specific memory traces fade over time. Also importantly, a view like theirs is not unique in the engram research literature. For example, Guskjolen and Cembrowski (2023) and Zaki and Cai (2025) hold similar views. They propose that highly specific hippocampal memory traces may decay rapidly, while more schematic and generalized neocortical information endures. If these views are correct, STM can accommodate the evidence in Section 4 by considering that recent memory traces are different from old memory traces. Consequently, they could argue that SH applies to remote memories, while CH applies only to recent ones.

However, the permanence of the identity of memory traces for particular events through consolidation processes remains highly controversial. What we can say is that the transience of memory traces, as predicted by the scene construction consolidation theory and similar views, is at least partially inconsistent with CH_2, and at the same time it helps SH accommodate the evidence discussed in Section 4 that was seen as inconsistent with simulationist views. Thus, STM, like CTM in the previous section, can rule out or explain away inconsistent evidence.

5.3 Co-allocation and generalizations

Sections 3 and 5.2 discussed the problem of generalizations for CTM, specifically because it is inconsistent with CH_1 and CH_2. In Section 5.2, I showed how consolidation processes can be seen as inconsistent with CTM because they lead to generalizations. In this subsection, I present a related process that can also lead to generalization and loss of identity of a memory trace by failing to represent the particular event involved in its formation (contra CH).

In co-allocation processes, information from different memory traces is integrated, resulting in a generalized representation (Richards & Frankland, 2017, Josselyn & Frankland, 2018). Many studies suggest that the memory system can maintain the identity of specific memory traces—that is, the memory trace retains its association with a specific event—even during co-allocation processes (e.g., Cai et al., 2016). However, some studies show that co-allocation can cause traces to lose their specific information (Rashid et al, 2016, Gastaldi et al, 2021). Andonovski's (2020, 2024) answer could be repeated here, stating that the integration of events via co-allocation would not be inconsistent with CH_1 and CH_2 because the system would still be manipulating representation of large yet particular events.

The problem remains because co-allocation is not restricted to temporally contiguous events (Richards & Frankland, 2017, Josselyn & Frankland, 2018). For example, Zaki and colleagues (2024) conducted a study in which fear responses generalized to different contexts after offline co-activation of engram ensembles during replay that resulted in co-allocation. Moreover, subtle similarities in contexts also lead to memory integration through co-allocation (Aly et al., 2022). In these studies, however, it is not clear whether the engrams have lost their identity. Given these cases, it would be counterintuitive to call them particular, unique events because they are not temporally continuous.

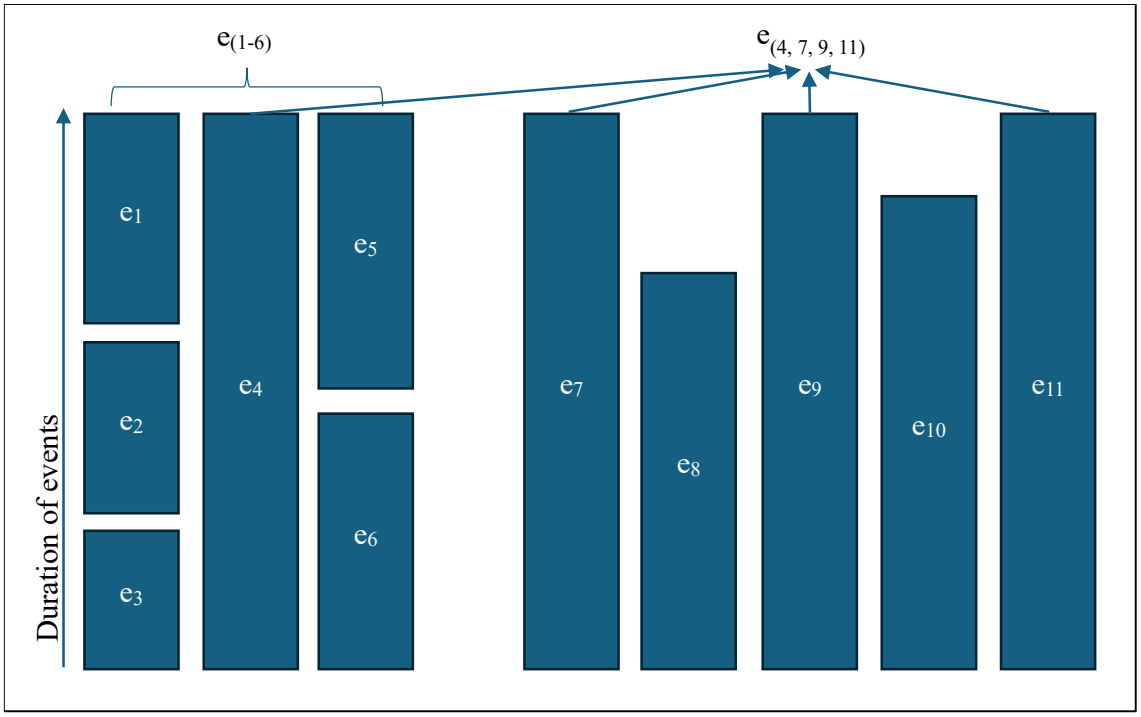

Fig. 1. Different types of individuation of events.

Many studies have shown that there are spatiotemporal and behavioral (goal-directed) conditions for the individuation of events (Franklin et al., 2020). This results in the same temporal duration being divided differently into individual events. For example, in Figure 1, $e_{(1-3)}$, e4, and $e_{(5-6)}$ have the same duration. However, as one can see in the figure, the same duration can be individuated differently into unique events (e_1-e_6 all have different durations). Given that events can be individuated differently with respect to duration, it is reasonable to consider $e_{(1-6)}$ (the integration of events e_1-e_6) as a particular, unique event, even if this integration occurs during consolidation rather than encoding,

because this integrated event refers to a unique spatiotemporal context (even if it is an extended spatiotemporal context).

Sometimes, however, non-continuous events, such as e_4, e_7, e_9, and e_{11}, are integrated not because of temporal proximity and continuity, but because of spatial and/or item similarity. As said, it is not clear that engrams lose their identity in these cases of integration. If they do lose their identity, it is odd to call $e_{(4,7,9,11)}$ a particular, unique event, because it does not refer to a particular, unique, spatiotemporal context or behavior, regardless of the time span of that context. $e_{(4,7,9,11)}$ seems to refer to a type rather than a token event, resulting in a loss of specificity (against CH_1). Crucially, at least some studies suggest that this type of generalization results in a loss of trace identity (against CH_2) (Rashid et al., 2016, Gastaldi et al., 2021).

If this is correct, the loss of trace identity through co-allocation would be only partially problematic for proponents of CH. While it would imply that traces can lose their identity—being partially inconsistent with CH_2—this constitutes weaker evidence than that from consolidation studies for two reasons. Firstly, the data supporting identity loss in co-allocation is more ambiguous, as the phenomenon is not yet firmly established. Secondly, even if proven, it would not affect all traces, but only a subset that undergoes co-allocation. Therefore, this evidence does not show that the episodic memory system never operates over traces of particular events; it only indicates that it *also* operates over some generalized traces. Consequently, evidence from co-allocation can, at best, suggest that CH is an incomplete characterization of the episodic memory system, not an incorrect one.

Causalists could also consider, in reply, that memory traces are processed by different systems. A semantic memory system would integrate them via co-allocation for generalizations, while the episodic memory system would ensure their specificity and process them individually. In this sense, causalists can point to studies showing that different immediate early genes are used to promote generalizations compared to those that ensure specificity (Nambu et al, 2022, Choucry et al, 2024, Yelhekar et al, 2024). That is, some traces' features are used to promote generalizations and have different functional and structural properties than traces used to ensure specificity. However, just as the loss of specificity in traces due to co-allocation is a hypothesis that needs further support, the functional differentiation of traces due to immediate early genes also needs further investigation.

In summary, the generalization properties of memory traces (either due to consolidation or due to co-allocation processes) may be problematic for CH_1-CH_2 because they suggest that specificity is not required for the functional profile of episodic memory. Importantly, as seen earlier, SH hypothesizes that the episodic system manipulates generalized representations to construct simulations. If the functional profile of the episodic system is to manipulate generalized information, then it is more consistent with STM than with CTM. For this reason, causalists still need to explain the generalization functions of memory traces if they are to defend CH_1-CH_2, as shown in Section 3. Proponents could, for instance, argue that while specific memory traces are necessary for the episodic memory system to function, they are not the only type of trace the system utilizes. Consequently, a more complete account of the system's functional profile would still be in the future for those who adopt CH.

Crucially, Section 4 presented evidence in favor of CH that not only supported CTM's hypotheses, but also provided a way for it to rule out some of the inconsistent data. Moreover, the evidence in Section 4 is seen as inconsistent with STM, which has led some philosophers to reject this latter theory of memory. However, as we will see in this section, CTM is also faced with inconsistent data, and there is further evidence involving memory traces that not only is compatible with SH, but also helps STM rule out some of its inconsistent data.

6 Conclusions

In this paper, I argued that the (anti)causalism debate is a debate about the functional profile of the episodic memory system. I also argued that the dispute is largely based on the functional properties of memory traces, i.e., what they represent, their discreteness, their formation, and their persistence. On the one hand, CTM defends that memory traces are discrete representations of particular events that are formed during the experience of those events and maintain their identity as representations of the same event over time. On the other hand, STM defends that the episodic system manipulates generalized information in the form of schemas and that memory traces, if they exist, either do not represent particular events or do not persist over time and maintain their identity.

Recently, some philosophers have returned to findings about the memory trace in order to defend CTM and reject STM. I showed, in light of the definition of CTM defended here, that the evidence they have appealed to is indeed incompatible with STM. However, I show that other evidence involving the functional properties of memory traces

is also inconsistent with CTM and helps STM resolve some of its problematic data. Because the evidence is thus inconclusive for either theory, I recommend caution in accepting or rejecting either theory of memory.

Importantly, the existence of inconsistent evidence for competing theories is not an anomaly in episodic memory research. In fact, as discussed in Section 2, even with inconsistent evidence, competing theories persist for a long time until the scientific community makes a decision to either reject or accept one of the competing theories. Crucially, inconsistent evidence for a particular theory should not be taken as a sufficient reason to reject it, because the inconsistency may not be due to deficiencies in the theory, but to deficiencies in the available data. In particular, in the case of memory traces and episodic memory, as argued by Barry and Maguire (2019a, 2019b), evidence for the maintenance of the identity of memory traces has not yet been found due to methodological limitations, as the studies are limited to one-month persistence, whereas some simulationist theories, such as theirs, predict that memory traces disappear after about four months. For this reason, more research on the functional nature of memory traces is needed to make a better decision about accepting/rejecting either CTM or STM.

References

Addis, D. R. (2020). Mental Time Travel? A Neurocognitive Model of Event Simulation. *Review of Philosophy and Psychology*, *11*(2), 233–259. https://doi.org/10.1007/s13164-020-00470-0

Allott, N., Lohndal, T., & Rey, G. (2021). Chomsky's "Galilean" Explanatory Style. In N. Allott, T. Lohndal, & G. Rey (Eds.), *A Companion to Chomsky* (1st ed., pp. 515–528). Wiley. https://doi.org/10.1002/9781119598732.ch33

Álvarez, J. F. (2025). Remembering and relearning: Against exclusionism. *Philosophical Studies*, *182*(2), 403–423. https://doi.org/10.1007/s11098-024-02265-w

Aly, M. H., Abdou, K., Okubo-Suzuki, R., Nomoto, M., & Inokuchi, K. (2022). Selective engram coreactivation in idling brain inspires implicit learning. *Proceedings of the National Academy of Sciences*, *119*(32), e2201578119. https://doi.org/10.1073/pnas.2201578119

Andonovski, N. (2020). SINGULARISM about Episodic Memory. *Review of Philosophy and Psychology*, *11*(2), 335–365. https://doi.org/10.1007/s13164-020-00464-y

Andonovski, N. (2021). Causation and mnemonic roles: On Fernández's Functionalism. *Estudios de Filosofía*, *64*, 139–153. https://doi.org/10.17533/udea.ef.n64a07

Andonovski, N. (2022). Episodic representation: A mental models account. *Frontiers in Psychology*, *13*, 899371. https://doi.org/10.3389/fpsyg.2022.899371

Andonovski, N. (2023). Autonoesis and the Galilean science of memory: Explanation, idealization, and the role of crucial data. *European Journal for Philosophy of Science*, *13*(3), 42. https://doi.org/10.1007/s13194-023-00548-3

Andonovski, N. (2024). Engrams as mental files. *Synthese*, *204*(6), 154. https://doi.org/10.1007/s11229-024-04801-x

Andonovski, N., & Michaelian, K. (2024). Naturalism and simulationism in the philosophy of memory. In Hossein Khani, A., Kemp, G., Rezaei, H. S., & Amiriara, H. (Eds.), *Naturalism and its challenges*. Routledge.

Andonovski, N., Sutton, J., & McCarroll, C. J. (2024). Eliminating episodic memory? *Philosophical Transactions of the Royal Society B: Biological Sciences*, *379*(1913), 20230413. https://doi.org/10.1098/rstb.2023.0413

Aronowitz, S. (forthcoming). Semanticization Challenges the Episodic–Semantic Distinction. *The British Journal for the Philosophy of Science*, 721760. https://doi.org/10.1086/721760

Asok, A., Leroy, F., Rayman, J. B., & Kandel, E. R. (2019). Molecular Mechanisms of the Memory Trace. *Trends in Neurosciences*, *42*(1), 14–22. https://doi.org/10.1016/j.tins.2018.10.005

Baker, B., Lansdell, B., & Kording, K. P. (2022). Three aspects of representation in neuroscience. *Trends in Cognitive Sciences*, *26*(11), 942–958. https://doi.org/10.1016/j.tics.2022.08.014

Barry, D. N., & Maguire, E. A. (2019a). Consolidating the Case for Transient Hippocampal Memory Traces. *Trends in Cognitive Sciences*, *23*(8), 635–636. https://doi.org/10.1016/j.tics.2019.05.008

Barry, D. N., & Maguire, E. A. (2019b). Remote Memory and the Hippocampus: A Constructive Critique. *Trends in Cognitive Sciences*, *23*(2), 128–142. https://doi.org/10.1016/j.tics.2018.11.005

Bergeron, V. (2007). Anatomical and Functional Modularity in Cognitive Science: Shifting the Focus. *Philosophical Psychology*, *20*(2), 175–195. https://doi.org/10.1080/09515080701197155

Bergeron, V. (2016). Functional Independence and Cognitive Architecture. *The British Journal for the Philosophy of Science*, *67*(3), 817–836. https://doi.org/10.1093/bjps/axv005

Bernecker, S. (2017). A Causal Theory of Mnemonic Confabulation. *Frontiers in Psychology*, *8*, 1207. https://doi.org/10.3389/fpsyg.2017.01207

Bernstein, D. M., & Loftus, E. F. (2009). How to Tell If a Particular Memory Is True or False. *Perspectives on Psychological Science*, *4*(4), 370–374. https://doi.org/10.1111/j.1745-6924.2009.01140.x

Bogen, J., & Woodward, J. (1988). Saving the Phenomena. *The Philosophical Review*, *97*(3), 303. https://doi.org/10.2307/2185445

Brunec, I. K., Robin, J., Olsen, R. K., Moscovitch, M., & Barense, M. D. (2020). Integration and differentiation of hippocampal memory traces. *Neuroscience & Biobehavioral Reviews*, *118*, 196–208. https://doi.org/10.1016/j.neubiorev.2020.07.024

Cai, D. J., Aharoni, D., Shuman, T., Shobe, J., Biane, J., Song, W., Wei, B., Veshkini, M., La-Vu, M., Lou, J., Flores, S. E., Kim, I., Sano, Y., Zhou, M., Baumgaertel, K., Lavi, A., Kamata, M., Tuszynski, M., Mayford, M., … Silva, A. J. (2016). A shared neural ensemble links distinct contextual memories encoded close in time. *Nature*, *534*(7605), 115–118. https://doi.org/10.1038/nature17955

Cao, R. (2022). Putting representations to use. *Synthese*, *200*(2), 151. https://doi.org/10.1007/s11229-022-03522-3

Chirimuuta, M. (2024). *The Brain Abstracted: Simplification in the History and Philosophy of Neuroscience*. The MIT Press. https://doi.org/10.7551/mitpress/13804.001.0001

Choucry, A., Nomoto, M., & Inokuchi, K. (2024). Engram mechanisms of memory linking and identity. *Nature Reviews Neuroscience*, *25*(6), 375–392. https://doi.org/10.1038/s41583-024-00814-0

Clewett, D., DuBrow, S., & Davachi, L. (2019). Transcending time in the brain: How event memories are constructed from experience. *Hippocampus*, *29*(3), 162–183. https://doi.org/10.1002/hipo.23074

Colaço, D. (2018). Rip it up and start again: The rejection of a characterization of a phenomenon. *Studies in History and Philosophy of Science Part A*, *72*, 32–40. https://doi.org/10.1016/j.shpsa.2018.04.003

Colaço, D., & Najenson, J. (2024). Where memory resides: Is there a rivalry between molecular and synaptic models of memory? *Philosophy of Science*, *91*(5), 1382–1392. https://doi.org/10.1017/psa.2023.126

Coltheart, M. (2011). Methods for modular modelling: Additive factors and cognitive neuropsychology. *Cognitive Neuropsychology*, *28*(3–4), 224–240. https://doi.org/10.1080/02643294.2011.587794

Coltheart, M. (2017). The assumptions of cognitive neuropsychology: Reflections on Caramazza (1984, 1986). *Cognitive Neuropsychology*, *34*(7–8), 397–402. https://doi.org/10.1080/02643294.2017.1324950

Craver, C. F. (2020). Remembering: Epistemic and Empirical. *Review of Philosophy and Psychology*, *11*(2), 261–281. https://doi.org/10.1007/s13164-020-00469-7

De Brigard, F. (2014). Is memory for remembering? Recollection as a form of episodic hypothetical thinking. *Synthese*, *191*(2), 155–185. https://doi.org/10.1007/s11229-013-0247-7

De Brigard, F. (2020). The Explanatory Indispensability of Memory Traces. *The Harvard Review of Philosophy*, *27*, 23–47. https://doi.org/10.5840/harvardreview202072328

De Brigard, F. (2023). *Memory and Remembering*. Cambridge University Press. https://doi.org/10.1017/9781108955447

De Brigard, F. (2024). Episodic memory without autonoetic consciousness. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *379*(1913), 20230410. https://doi.org/10.1098/rstb.2023.0410

De Brigard, F. (forthcoming). Simulationism and memory traces. In S. Aranowitz, & L. Nadel. *Memory, Space, and Time.* Oxford University Press.

De Cruz, H., & De Smedt, J. (2013). The value of epistemic disagreement in scientific practice. The case of Homo floresiensis. *Studies in History and Philosophy of Science Part A*, *44*(2), 169–177. https://doi.org/10.1016/j.shpsa.2013.02.002

Diamond, N. B., Armson, M. J., & Levine, B. (2020). The Truth Is Out There: Accuracy in Recall of Verifiable Real-World Events. *Psychological Science*, *31*(12), 1544–1556. https://doi.org/10.1177/0956797620954812

Ekstrom, A. D., & Yonelinas, A. P. (2020). Precision, binding, and the hippocampus: Precisely what are we talking about? *Neuropsychologia*, *138*, 107341. https://doi.org/10.1016/j.neuropsychologia.2020.107341

Finley, J. R., & Brewer, W. F. (2024). Accuracy and completeness of autobiographical memory: Evidence from a wearable camera study. *Memory*, *32*(8), 1012–1042. https://doi.org/10.1080/09658211.2024.2377193

Foudil, S.-A., Pleche, C., & Macaluso, E. (2021). Memory for spatio-temporal contextual details during the retrieval of naturalistic episodes. *Scientific Reports*, *11*(1), 14577. https://doi.org/10.1038/s41598-021-93960-9

Franklin, N. T., Norman, K. A., Ranganath, C., Zacks, J. M., & Gershman, S. J. (2020). Structured Event Memory: A neuro-symbolic model of event cognition. *Psychological Review*, *127*(3), 327–361. https://doi.org/10.1037/rev0000177

Friston, K. J., & Price, C. J. (2011). Modules and brain mapping. *Cognitive Neuropsychology*, *28*(3–4), 241–250. https://doi.org/10.1080/02643294.2011.558835

Gastaldi, C., Schwalger, T., De Falco, E., Quiroga, R. Q., & Gerstner, W. (2021). When shared concept cells support associations: Theory of overlapping memory engrams. *PLOS Computational Biology*, *17*(12), e1009691. https://doi.org/10.1371/journal.pcbi.1009691

Gershman, S. J. (2023). The molecular memory code and synaptic plasticity: A synthesis. *Biosystems*, *224*, 104825. https://doi.org/10.1016/j.biosystems.2022.104825

Gershman, S. J., Radulescu, A., Norman, K. A., & Niv, Y. (2014). Statistical Computations Underlying the Dynamics of Memory Updating. *PLoS Computational Biology*, *10*(11), e1003939. https://doi.org/10.1371/journal.pcbi.1003939

Guskjolen, A., & Cembrowski, M. S. (2023). Engram neurons: Encoding, consolidation, retrieval, and forgetting of memory. *Molecular Psychiatry*, *28*(8), 3207–3219. https://doi.org/10.1038/s41380-023-02137-5

Hassabis, D., & Maguire, E. A. (2009). The construction system of the brain. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1521), 1263–1271. https://doi.org/10.1098/rstb.2008.0296

Heinen, R., Bierbrauer, A., Wolf, O. T., & Axmacher, N. (2024). Representational formats of human memory traces. *Brain Structure and Function*, *229*(3), 513–529. https://doi.org/10.1007/s00429-023-02636-9

Josselyn, S. A., & Frankland, P. W. (2018). Memory Allocation: Mechanisms and Function. *Annual Review of Neuroscience*, *41*(1), 389–413. https://doi.org/10.1146/annurev-neuro-080317-061956

Josselyn, S. A., & Tonegawa, S. (2020). Memory engrams: Recalling the past and imagining the future. *Science*, *367*(6473), eaaw4325. https://doi.org/10.1126/science.aaw4325

Kitcher, P. (1990). The Division of Cognitive Labor. *The Journal of Philosophy*, *87*(1), 5. https://doi.org/10.2307/2026796

Klein, C. (2010). Images Are Not the Evidence in Neuroimaging. *The British Journal for the Philosophy of Science*, *61*(2), 265–278. https://doi.org/10.1093/bjps/axp035

Kunz, L., Deuker, L., Zhang, H., & Axmacher, N. (2018). Tracking Human Engrams Using Multivariate Analysis Techniques. In *Handbook of Behavioral Neuroscience* (Vol. 28, pp. 481–508). Elsevier. https://doi.org/10.1016/B978-0-12-812028-6.00026-4

Langland-Hassan, P. (2022). Propping up the causal theory. *Synthese*, *200*(2), 95. https://doi.org/10.1007/s11229-022-03635-9

Loftus, E. F., Levidow, B., & Duensing, S. (1992). Who remembers best? Individual differences in memory for events that occurred in a science museum. *Applied Cognitive Psychology*, *6*(2), 93–107. https://doi.org/10.1002/acp.2350060202

Luis-Islas, J., Luna, M., Floran, B., & Gutierrez, R. (2022). Optoception: Perception of Optogenetic Brain Perturbations. *Eneuro*, *9*(3), ENEURO.0216-22.2022. https://doi.org/10.1523/ENEURO.0216-22.2022

Mahon, B. Z., & Cantlon, J. F. (2011). The specialization of function: Cognitive and neural perspectives. *Cognitive Neuropsychology*, *28*(3–4), 147–155. https://doi.org/10.1080/02643294.2011.633504

Martin, C. B., & Deutscher, M. (1966). Remembering. *The Philosophical Review*, *75*(2), 161. https://doi.org/10.2307/2183082

McCarroll, C. J. (2023). Memory and Imagination, Minds and Worlds. In A. Berninger & Í. Vendrell Ferran, *Philosophical Perspectives on Memory and Imagination* (pp. 35–53). Routledge. https://doi.org/10.4324/9781003153429-4

McCarroll, C. J., Michaelian, K., & Nanay, B. (2024). Explanatory Contextualism about Episodic Memory: Towards A Diagnosis of the Causalist-Simulationist Debate. *Erkenntnis*, *89*(6), 2273–2301. https://doi.org/10.1007/s10670-022-00629-4

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*(3), 419–457. https://doi.org/10.1037/0033-295X.102.3.419

Ménager, D. H., Choi, D., & Robins, S. K. (2021). A Hybrid Theory of Event Memory. *Minds and Machines*, *32*(2), 365–394. https://doi.org/10.1007/s11023-021-09578-3

Ménager, D. H., Choi, D., & Robins, S. K. (2022). Modeling human memory phenomena in a hybrid event memory system. *Cognitive Systems Research*, *75*, 25–35. https://doi.org/10.1016/j.cogsys.2022.05.003

Michaelian, K. (2016). *Mental time travel: Episodic memory and our knowledge of the personal past*. MIT Press.

Michaelian, K. (2021). Imagining the past reliably and unreliably: Towards a virtue theory of memory. *Synthese*, *199*(3–4), 7477–7507. https://doi.org/10.1007/s11229-021-03125-4

Michaelian, K. (2022). Against Perrin's embodied causalism: Still no evidence for the necessity of appropriate causation. *Intellectica, 76*, 175-190.

Michaelian, K. (2024). Radicalizing simulationism: Remembering as imagining the (nonpersonal) past. *Philosophical Psychology*, 1–27. https://doi.org/10.1080/09515089.2022.2082934

Michaelian, K., Álvarez, J.F., & Openshaw, J. (forthcoming). Is De Brigard a simulationist? *Philosophy and the Mind Sciences*.

Michaelian, K., & Robins, S. K. (2018). Beyond the causal theory? Fifty years after Martin and Deutscher. In K. Michaelian, D. Debus, & D. Perrin (Eds.), *New directions in the philosophy of memory* (pp. 13–32). Routledge, Taylor & Francis Group.

Morales, C., Morici, J. F., Miranda, M., Gallo, F. T., Bekinschtein, P., & Weisstaub, N. V. (2020). Neurophotonics Approaches for the Study of Pattern Separation. *Frontiers in Neural Circuits*, *14*, 26. https://doi.org/10.3389/fncir.2020.00026

Moscovitch, M., & Gilboa, A. (2024). Systems Consolidation, Transformation, and Reorganization: Multiple Trace Theory, Trace Transformation Theory, and Their Competitors. In M. J. Kahana & A. D. Wagner (Eds.), *The Oxford Handbook of Human Memory, Two Volume Pack* (1st ed., pp. 1278–1328). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190917982.013.43

Moscovitch, M., & Nadel, L. (2019). Sculpting Remote Memory: Enduring Hippocampal Traces and vmPFC Reconstructive Processes. *Trends in Cognitive Sciences*, *23*(8), 634–635. https://doi.org/10.1016/j.tics.2019.05.001

Mullally, S. L., & Maguire, E. A. (2014). Memory, Imagination, and Predicting the Future: A Common Brain Mechanism? *The Neuroscientist*, *20*(3), 220–234. https://doi.org/10.1177/1073858413495091

Najenson, J. (2021). What have we learned about the engram? *Synthese*, *199*(3–4), 9581–9601. https://doi.org/10.1007/s11229-021-03216-2

Najenson, J. (forthcoming). Encoding without perceiving: Can memories be implanted? *Philosophical Psychology*, 1–28. https://doi.org/10.1080/09515089.2023.2295927

Nambu, M. F., Lin, Y.-J., Reuschenbach, J., & Tanaka, K. Z. (2022). What does engram encode?: Heterogeneous memory engrams for different aspects of experience. *Current Opinion in Neurobiology*, *75*, 102568. https://doi.org/10.1016/j.conb.2022.102568

Openshaw, J. (2023). (In defence of) preservationism and the previous awareness condition: What is a theory of remembering, anyway? *Philosophical Perspectives*, *37*(1), 290–307. https://doi.org/10.1111/phpe.12191

Openshaw, J., & Michaelian, K. (2024). Reference in remembering: Towards a simulationist account. *Synthese*, *203*(3), 90. https://doi.org/10.1007/s11229-024-04508-z

Park, S., Ko, S. Y., Frankland, P. W., & Josselyn, S. A. (2024). Comparing behaviours induced by natural memory retrieval and optogenetic reactivation of an engram ensemble in mice. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *379*(1906), 20230227. https://doi.org/10.1098/rstb.2023.0227

Parkkinen, V.-P., Russo, F. & Wallmann, C. (2017). Scientific disagreement and evidential pluralism: lessons from the studies on hypercholesterolemia. *HUMANA.MENTE: Journal of Philosophical Studies*, 32:75-116.

Piccinini, G. (2022). Situated Neural Representations: Solving the Problems of Content. *Frontiers in Neurorobotics*, *16*, 846979. https://doi.org/10.3389/fnbot.2022.846979

Pietroski, P., & Rey, G. (1995). When Other Things Aren't Equal: Saving *Ceteris Paribus* Laws from Vacuity. *The British Journal for the Philosophy of Science*, *46*(1), 81–110. https://doi.org/10.1093/bjps/46.1.81

Potochnik, A. (2017). *Idealization and the aims of science*. The University of Chicago press.

Powers, P. A., Andriks, J. L., & Loftus, E. F. (1979). Eyewitness accounts of females and males. *Journal of Applied Psychology*, *64*(3), 339–347. https://doi.org/10.1037/0021-9010.64.3.339

Ranganath, C. (2024). Episodic Memory. In M. J. Kahana & A. D. Wagner (Eds.), *The Oxford Handbook of Human Memory, Two Volume Pack* (1st ed., pp. 151–171). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190917982.013.6

Ranganath, C., & Ritchey, M. (2012). Two cortical systems for memory-guided behaviour. *Nature Reviews Neuroscience*, *13*(10), 713–726. https://doi.org/10.1038/nrn3338

Rashid, A. J., Yan, C., Mercaldo, V., Hsiang, H.-L. (Liz), Park, S., Cole, C. J., De Cristofaro, A., Yu, J., Ramakrishnan, C., Lee, S. Y., Deisseroth, K., Frankland, P. W., & Josselyn, S. A. (2016a). Competition between engrams influences fear memory formation and recall. *Science*, *353*(6297), 383–387. https://doi.org/10.1126/science.aaf0594

Richards, B. A., & Frankland, P. W. (2017). The Persistence and Transience of Memory. *Neuron*, *94*(6), 1071–1084. https://doi.org/10.1016/j.neuron.2017.04.037

Ritchey, M., & Cooper, R. A. (2020). Deconstructing the Posterior Medial Episodic Network. *Trends in Cognitive Sciences*, *24*(6), 451–465. https://doi.org/10.1016/j.tics.2020.03.006

Ritchey, M., Libby, L. A., & Ranganath, C. (2015). Cortico-hippocampal systems involved in memory and cognition. In *Progress in Brain Research* (Vol. 219, pp. 45–64). Elsevier. https://doi.org/10.1016/bs.pbr.2015.04.001

Robins, S. K. (2016). Representing the past: Memory traces and the causal theory of memory. *Philosophical Studies*, *173*(11), 2993–3013. https://doi.org/10.1007/s11098-016-0647-x

Robins, S. K. (2017). Contiguity and the causal theory of memory. *Canadian Journal of Philosophy*, *47*(1), 1–19. https://doi.org/10.1080/00455091.2016.1209964

Robins, S. K. (2020). Stable Engrams and Neural Dynamics. *Philosophy of Science*, *87*(5), 1130–1139. https://doi.org/10.1086/710624

Robins, S. K. (2022). The Role of Memory Science in the Philosophy of Memory. *Philosophy Compass*, *17*(10), e12880. https://doi.org/10.1111/phc3.12880

Robins, S. K. (2023). The 21st century engram. *WIREs Cognitive Science*, *14*(5), e1653. https://doi.org/10.1002/wcs.1653

Robins, S. K. (forthcoming). A place for the memory trace. In Aronowitz, S. & Nadel, L. (eds.). *Memory, Space, and Time*. Oxford University Press.

Rosenbaum, R. S., Köhler, S., Schacter, D. L., Moscovitch, M., Westmacott, R., Black, S. E., Gao, F., & Tulving, E. (2005). The case of K.C.: Contributions of a memory-impaired person to memory theory. *Neuropsychologia*, *43*(7), 989–1021. https://doi.org/10.1016/j.neuropsychologia.2004.10.007

Roskies, A. L. (2021). Representational similarity analysis in neuroimaging: Proxy vehicles and provisional representations. *Synthese*, *199*(3–4), 5917–5935. https://doi.org/10.1007/s11229-021-03052-4

Santana, C. (2021). Let's not agree to disagree: The role of strategic disagreement in science. *Synthese*, *198*(S25), 6159–6177. https://doi.org/10.1007/s11229-019-02202-z

Sant'Anna, A. (2021). Mnemonic causation, construction, and the particularity of episodic memory. *Aufklärung: Journal of Philosophy*, *8*(esp), 57–70. https://doi.org/10.18012/arf.v8iesp.60017

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. https://doi.org/10.1098/rstb.2007.2087

Schacter, D. L., Addis, D. R., Hassabis, D., Martin, V. C., Spreng, R. N., & Szpunar, K. K. (2012). The Future of Memory: Remembering, Imagining, and the Brain. *Neuron*, *76*(4), 677–694. https://doi.org/10.1016/j.neuron.2012.11.001

Schacter, D. L., Guerin, S. A., & St. Jacques, P. L. (2011). Memory distortion: An adaptive perspective. *Trends in Cognitive Sciences*, *15*(10), 467–474. https://doi.org/10.1016/j.tics.2011.08.004

Schacter, D. L., & Tulving, E. (1994). What are the memory systems of 1994? In D. L. Schacter & E. Tulving (Eds.), *Memory systems 1994.* The MIT Press.

Schirmer dos Santos, C., McCarroll, C. J., & Sant'Anna, A. (2022). The relation between memory and imagination: A debate about the right concepts. In A. Sant'Anna, C. J. McCarroll, & K. Michaelian (Eds.), *Current controversies in philosophy of memory* (pp. 38–56). Routledge.

Shallice, T. (1988). *From neuropsychology to mental structure*. Cambridge University Press.

Siegel, G., & Craver, C. F. (2024). Phenomenological Laws and Mechanistic Explanations. *Philosophy of Science*, *91*(1), 132–150. https://doi.org/10.1017/psa.2023.65

Squire, L. R., Genzel, L., Wixted, J. T., & Morris, R. G. (2015). Memory Consolidation. *Cold Spring Harbor Perspectives in Biology*, *7*(8), a021766. https://doi.org/10.1101/cshperspect.a021766

Steel, D. P. (2008). *Across the boundaries: Extrapolation in biology and social science*. Oxford University Press.

Sternberg, S. (2011). Modular processes in mind and brain. *Cognitive Neuropsychology*, *28*(3–4), 156–208. https://doi.org/10.1080/02643294.2011.557231

Theves, S., Grande, X., Duzel, E., & Doeller, C. F. (2024). Pattern Completion and the Medial Temporal Lobe Memory System. In M. J. Kahana & A. D. Wagner (Eds.),

The Oxford Handbook of Human Memory, Two Volume Pack (1st ed., pp. 988–1016). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190917982.013.34

Tonegawa, S., Morrissey, M. D., & Kitamura, T. (2018). The role of engram cells in the systems consolidation of memory. *Nature Reviews Neuroscience*, *19*(8), 485–498. https://doi.org/10.1038/s41583-018-0031-2

Tulving, E. (1972). Episodic and Semantic Memory. In E. Tulving, W. Donaldson, G. H. Bower, & United States (Eds.), *Organization of memory*. Academic Press.

Tulving, E. (1983). *Elements of episodic memory*. Clarendon Press.

Tulving, E. (1984). How many memory systems are there? *American Psychologist, 40*(4), 385.

Tulving, E. (1986). What kind of a hypothesis is the distinction between episodic and semantic memory? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *12*(2), 307–311. https://doi.org/10.1037/0278-7393.12.2.307

Tulving, E. (1999). Study of memory: Processes and systems. In J. K. Foster & M. E. Jelicic (Eds.), *Memory: Systems, process, or function?* Oxford University Press.

Memory: Systems, process, or function? Oxford University Press.

Tulving, E. (2002). Episodic Memory: From Mind to Brain. *Annual Review of Psychology*, *53*(1), 1–25. https://doi.org/10.1146/annurev.psych.53.100901.135114

Vecchi, T., & Gatti, D. (2020). *Memory as prediction: From looking back to looking forward*. The MIT press.

Vetere, G., Tran, L. M., Moberg, S., Steadman, P. E., Restivo, L., Morrison, F. G., Ressler, K. J., Josselyn, S. A., & Frankland, P. W. (2019). Memory formation in the absence of experience. *Nature Neuroscience*, *22*(6), 933–940. https://doi.org/10.1038/s41593-019-0389-0

Vilarroya, O. (2017). Neural Representation. A Survey-Based Analysis of the Notion. *Frontiers in Psychology*, *8*, 1458. https://doi.org/10.3389/fpsyg.2017.01458

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. https://doi.org/10.1007/s13164-020-00471-z

Woodward, J. F. (2011). Data and phenomena: A restatement and defense. *Synthese*, *182*(1), 165–179. https://doi.org/10.1007/s11229-009-9618-5

Xue, G. (2022). From remembering to reconstruction: The transformative neural representation of episodic memory. *Progress in Neurobiology*, *219*, 102351. https://doi.org/10.1016/j.pneurobio.2022.102351

Yelhekar, T. D., Meng, M., Doupe, J., & Lin, Y. (2024). All IEGs Are Not Created Equal—Molecular Sorting Within the Memory Engram. In J. Gräff & S. Ramirez (Eds.), *Engrams* (Vol. 38, pp. 81–109). Springer International Publishing. https://doi.org/10.1007/978-3-031-62983-9_6

Yonelinas, A. P., Ranganath, C., Ekstrom, A. D., & Wiltgen, B. J. (2019). A contextual binding theory of episodic memory: Systems consolidation reconsidered. *Nature Reviews Neuroscience*, *20*(6), 364–375. https://doi.org/10.1038/s41583-019-0150-4

Zaki, Y., & Cai, D. J. (2025). Memory engram stability and flexibility. *Neuropsychopharmacology*, *50*(1), 285–293. https://doi.org/10.1038/s41386-024-01979-z

Zaki, Y., Pennington, Z. T., Morales-Rodriguez, D., Bacon, M. E., Ko, B., Francisco, T. R., LaBanca, A. R., Sompolpong, P., Dong, Z., Lamsifer, S., Chen, H.-T., Carrillo Segura, S., Christenson Wick, Z., Silva, A. J., Rajan, K., Van Der Meer, M., Fenton, A., Shuman, T., & Cai, D. J. (2024). Offline ensemble co-reactivation links memories across days. *Nature*. https://doi.org/10.1038/s41586-024-08168-4

Zollman, K. J. S. (2010). The Epistemic Benefit of Transient Diversity. *Erkenntnis*, *72*(1), 17–35. https://doi.org/10.1007/s10670-009-9194-6