1. Introduction

A key debate in the philosophy of memory concerns the nature of remembering: what is required for $S$ to remember $p$? The standard view, for several decades now, has been the Causal Theory of Memory (e.g., Martin & Deutscher, 1966; Bernecker, 2010; Debus, 2010). According to the Causal Theory, remembering requires a particular kind of causal connection between the past event and subsequent recall—i.e., a connection sustained by a memory trace. Simulationism (Michaelian 2016a; 2016b) offers a recent and increasingly popular alternative account of remembering. For Simulationists, the memory trace is replaced with a reliable memory system.

The appeal to reliability for assessments of remembering seems straightforward enough. After all, epistemologists have already spent a lot of time thinking about the reliability of cognitive process in general—and of memory in particular. I won’t have anything to say about this work today (and, at least in broad strokes, I am on board with this project). My argument instead is focused on the exportation of that notion of reliability from this original context into debates about the nature of remembering. I will argue that the move simply doesn’t work; the notion of reliability has no substantive role to play in an account of remembering. I attempt to motivate this idea by showing, first, the differences between the investigatory projects in the epistemology of memory and debates about remembering. I then turn to a direct comparison of my version of the Causal Theory and Michaelian’s Simulationism, arguing that his account fails to render the correct judgment on several key cases—and that the failure is due to the reliance on reliability. I conclude with two general claims about the incompatibility of remembering and reliability.

2. Preliminaries: Two Key Contrasts

My claim that there is no significant role for the notion to play in assessments of remembering may strike the reader as odd. The Simulationist move looks like a mere extension of a fruitful and widely
accepted framework. There are, however, subtle but important differences in how memory is understood and how it relates to the central questions being asked in each domain. My aim in this section is to draw out these differences and create some space between the questions about memory that have been of primary concern in epistemology and the questions about memory that are at the forefront of debates in the philosophy of memory.

In epistemology, the central question (I take it) is this: Is memory reliable? Some arrive at this question from particular, reliabilist views about justification, but the question also arises more generally as part of a broader concern about the possibility of memory knowledge. Tom Senor illustrates this latter point nicely in the introduction to his SEP entry on the Epistemology of Memory:

That most of our knowledge is in memory at any particular time is a given. What is perhaps surprising, however, is the degree to which even our current conscious knowledge typically depends on memory. For example, you look at the sky and come to believe that the sunset is beautiful. This is a newly formed belief about an event currently taking place. Nevertheless, its justification is no doubt dependent on other beliefs that you hold. For example, if you didn't at least tacitly believe that you were looking west or that it is evening and not morning, the belief wouldn't be justified (I assume that the phenomenology of sunsets and sunrises is indistinguishable). Now I am not, however, supposing that all knowledge of the external world is inferential. Your belief that the sky is red might well be epistemically basic. Nor am I insisting that your sunset belief is psychologically inferential. My only claim here is that many relatively simple beliefs we form about the external world typically depend for their justification on background beliefs; and background beliefs are memory beliefs (Senor 2009: p. 1)

If memory beliefs constitute a large proportion of our total beliefs that are candidates for knowledge, then it becomes crucial that the memory system which stores/produces these beliefs is reliable (even if one does not consider memory’s reliability to be the source of their justification).

Epistemological concerns with memory are thus concerns about how memory influences our ability to hold true or mostly-true beliefs—it’s role in storing, producing, and/or justifying those beliefs. This focus is understandable, given the kinds of questions epistemologists want to answer. The motivations for developing accounts of remembering in philosophy of memory are different, and so call upon a distinct sense of memory and do so in service of different questions. In what follows, I identify two key differences in how memory is understood and used.

---

1 In fact, insofar as Goldman’s (1975) initial formulations of reliabilism were an attempt to refine the causal theory of knowing one might see the simulationist move to refine the causal theory of memory as a similar step in the right direction.
2.1 Semantic vs. Episodic

Although both may involve memory, the projects in epistemology and philosophy of memory involve two distinct forms of memory. As the Senor quote above makes clear, the primary epistemological concern is with memory beliefs. Memory beliefs are the products of semantic memory. Philosophers of memory are concerned primarily with episodic memory. The distinction between semantic and episodic memory is, roughly, a distinct between memory for facts and memory for events or experiences. Remembering that Dalmatians are a breed of dog is a semantic memory; remembering the day I adopted a Dalmatian puppy named Zippy is an episodic memory. This distinction is generally endorsed by both philosophers and memory scientists, although in philosophy it is sometimes characterized as a distinction between propositional and experiential memory, respectively. It can be difficult to make the distinction precise, but it is a distinction that nearly everyone who wants to characterize memory wants to make.

As a general rule, most philosophers inquiring into the nature of remembering are focused on episodic memory. Episodic memory is memory for particular past experiences. They are memories of past events that the rememberer herself experienced. They include both the mundane and the momentous—where one’s car is parked in the lot and the birth of one’s first child. Individual accounts may differ in how they characterize episodic memory—e.g., whether it is thought to involve mental time travel or other signature phenomenological features; whether it is exclusively human or shared with other animal species; how it relates to our ability to imagine ourselves in the future or in possible, but not actual pasts. Philosophers of memory do not spend much time thinking (or at least writing) about semantic memory. The interest is in episodic memory—and for its own sake, rather than in comparison to the semantic.

The focus is understandable. Given that philosophers of memory (like myself) have a primary interest in memory – as a psychological/cognitive capacity or ability – then the inclination is

---

2 In psychology, the episodic/semantic distinction comes from the work of Endel Tulving. Tulving has devoted much of his career to refining the distinction between semantic and episodic memory. In 1972, Tulving claimed that episodic memory involved spatial and temporal information, made autobiographical reference to the self, and was more susceptible to forgetting than semantic memory. In 1982, he added that the two systems were functionally separable, and in 1985 added a further claim of stochastic independence. Tulving (2002) has since rejected the claims of functional and stochastic independence, replacing them with a sense of subjective time, autonoetic awareness, and a concept of the self.
to evaluate and consider it at its most distinctive. And episodic remembering is arguably the most paradigmatic form of memory. If one were to ask you to report a memory, I would bet money that you’d select an anecdote from childhood (or momentous occasion) rather than a general fact. You’d be more likely to tell me about the time you made it backstage at a Prince concert than that the population of Omaha is 466,000 (unless of course the population is somehow relevant to the story—the kind of intertwining that makes distinguishing these forms of memory complicated). Since philosophers of memory are interested in saying what is distinctive and characteristic of memory, their interests go towards its emblematic, episodic form. Focusing here helps to highlight what’s interesting about memory as memory, at its most distinguishable from other mental states (like perception and belief). The reason is, again, understandable—and inclines in the opposite direction of the interests in epistemology (where it is the similarity between stored belief and other forms of belief that brings memory into the discussion).

Pointing this out serves as more than a sociological point. The difference in focus brings along further differences in the requirements or restrictions on the mental states under consideration. As forms of memory, semantic and episodic memory share a diachronic nature: they are mental states/capacities/abilities that are extended over time. Or, put another way, considering an instance of semantic or episodic remembering as remembering entails that there was a prior event when the information, experience, etc. was acquired. But there are differences in the strength of this diachronic requirement. Consider the two Dalmatian-involving memories from earlier. The former, a general fact about Dalmatians, could be acquired in a number of different ways: by reading a book, watching a documentary, being raised by parents who were dog breeders, etc. In order to be an instance of memory, there must be a connection to some previous event during which the information was required, but no connection to any particular previous event is required. The memory of adopting a Dalmatian puppy, in contrast, requires a connection to a very particular past event—namely, the event where I adopted Zippy. When our concern is episodic memory, our interest is in a particular past event.

2.2 Content vs. Attitude
The concern about memory beliefs, in epistemology, is with whether they are true (or mostly true). When philosophers of memory are giving an account of remembering, the focus is not on the truth, but on the status of the mental state as a memory. This is not to say that truth is irrelevant, of course. Truth is involved, insofar as most accounts of memory assume that memory is factive, or at least
that it has to be by and large accurate (more on that in the next section). But the question at issue for philosophers of memory is about what requirements over and above accuracy are required for remembering. This influences the way in which reliability can enter into the two projects. For knowledge questions, I am asking whether memory produces true beliefs reliably. For remembering questions, I am asking whether the state produced is, reliably, one of remembering. The focus is on whether the way in which one came to hold that (true) belief counts as an instance of remembering.

To see the point, let’s continue with the example of my adopting Zippy. The question is not whether my memory of adopting Zippy is true of the past event—that is a requirement on remembering, but meeting this requirement alone isn’t sufficient for remembering. Instead, it’s a question of whether the way I’ve come to be in that mental state (which may be a state of belief or a state of seeming to remember, etc – this is a point of contention) is constitutive of remembering. After all, there are many ways that I could come to hold true beliefs about my past experiences. I could, for example, be led to form beliefs about this past experience inferentially. I could know that the first dog my family ever adopted was named Zippy and that my family had a strong preference for adopting Dalmatians…and so on. Or my representation of the event could be based on information I later acquired about the event, rather than on the event itself. Childhood ‘memories’ are the best illustration of this. Maybe the adoption of Zippy happened when I was a young, and is a story that my family has enjoyed re-telling at annual gatherings ever since. Although I may have a very vivid and (let’s suppose) fully accurate representation of this past event—one that feels very much like a memory—it’s still reasonable for me to wonder whether I actually remember adopting Zippy or only feel this way as a result of the repetitive family re-hashings. Depending on my inferential skills and the truthfulness of my family’s storytelling, these ways of forming beliefs about my particular past event may be true, and even reliably so. But they’re not ways of remembering the experience.3

This further separates the epistemological and philosophy of memory projects. The question about requirements for remembering concerns not just mental states about particular past

---

3 It’s difficult to formulate this point in a way that resists begging the question against the Simulationist (or other non-causalist), since many of those views will allow a wider range of states to count as remembering than is allowed by the Causalist. But even if the Simulationist would want to allow some of the cases I’ve mentioned here into the account of remembering, it’s safe to say that they would not want to allow in all true beliefs about past events and experiences. They will want to draw the line somewhere and, so long as that’s the case, my point about the distance between the two projects stands.
events, but mental states about particular past events that are formed and maintained in a particular way. Exactly how this particular way should be understood is the point of disagreement between Causalist and Simulationist accounts of remembering. Even if the argument I’ll go on to make (i.e., that the Simulationist account of remembering fails, in large part because of its reliance on reliability) is unsuccessful, I hope to have at least established that the way in which reliability is imported into debates about remembering should be attended to and treated with caution, as the concept will have different work to do in this context than it does in conversations about the possibility of memory knowledge.

3. Remembering: Causalist vs. Simulationist Approaches

3.1 Preliminaries
For the purposes of this paper, I’m framing the debate over the nature of remembering as one of competing accounts of what is required for S to remember p. In focusing on S remembers p (rather than S remembers that p)—I’m trying to both signal the difference from questions about semantic memory that were highlighted in the previous section and also attempting to avoid getting bogged down in the differences between individual accounts of how the forms of memory are characterized and classified.

In setting the requirements for S remembers p, Causalists and Simulationists don’t disagree about everything—in fact, they agree about two of the three requirements on remembering. First, is an accuracy condition. In order for S to remember p, p must be the case (or be true, or S’s representation must accurately depict, resemble, or be made true by p). Sometimes this point is made by claiming that memory is factive—a success term, which builds accurate remembering into the account of remembering itself. There are plenty of worries one can have about accuracy assessments—whether there can be updates or changes to the memory over time, and if so, how much—but in facing these, the Causalist and Simulationist are fighting on the same team (or at least facing the same challengers). Second, there is an identity condition. Remembering is a diachronic capacity, requiring mental activity at (at least) two distinct times: the initial event, where the information is acquired, and the subsequent event where it is remembered. The identity condition stipulates that the person involved in the initial event is numerically identical to the person involved in the subsequent event. Despite its rejection of a common way of thinking about the relation and identity amongst philosophers working on issues of personal identity, this is not a point that most
philosophers of memory go to great lengths to defend. Instead, it is often simply built into the account—for example, by giving an account of S remembers p, where there are appeals to features of S at both t₁ and t₂ (e.g., Michaelian 2011). As with the accuracy condition, I am not presenting this claim as beyond reproach or without complications, but merely setting it aside as a condition which will unite rather than divide Causalists and Simulationists.

The condition that does divide Causalists and Simulationists concerns what is required for remembering over and above accuracy and identity. There are likely many ways of accurately representing information about one’s own past; remembering is a proper subset of the states that meet these first two conditions. Causalists argue that this ‘something more’ is a causal connection between the initial event and the subsequent remembering, but not just any causal connection will do: it requires a connection via a memory trace. Simulationists reject the appeal to a memory trace in an account of remembering, generally on the grounds that such a commitment is incompatible with memory’s reconstructive character. In place of the trace, Simulationists appeal to a reliable memory system.

Debate between Causal and Simulationist views is carried out largely by consideration of how well each account does at accounting for memory errors and at distinguishing these errors from each other and from successful remembering. The errors used in this process are all errors of commission rather than omission (i.e., forgetting). There are three commission errors on which most attention has been focused: misremembering, confabulation, and relearning. Introducing them outside of any particular theory is difficult because Causalists and Simulationists differ not only in how they explain these errors, but in how they define them. I will, however, attempt a general sketch prior to introducing either view.

Misremembering and confabulation are errors of memory that are generally understood as analogous to illusion and hallucination, respectively, as errors of perception. That is, misremembering is understood as a partial error—one where the memory is distorted or incorrect in

---

4 This identity claim is often rejected by philosophers working on issues of personal identity, who may want to use memory as the basis for personal identity and cannot do so if memory is defined in terms of identity, on pain of circularity. Doing so requires a notion of quasi-memory, which many in philosophy of memory do not consider a viable psychological notion (see Schectman 1990 for an argument on this point).

5 It’s worth noting that there are other accounts of remembering where nothing more may be required – e.g., the Simple Retention View (as labeled by Bernecker 2010; attributed to Ryle 1949, Squires 1969, etc.)
some aspects but not in others. Confabulation, on the other hand, is a total error, lacking the necessary feature or features entirely. Relearning errors are particular to memory, and considered by some to be particular to the Causal Theory of Memory, as they were introduced by Martin and Deutscher (1966). Relearning occurs when someone learns something, forgets it entirely, and then learns it again. Childhood ‘memories,’ like the adoption of Zippy from above, are illustrative examples of relearning. In such cases, the mental state is accurate and involves the same person who experienced the event (and so meets the first two requirements on remembering). The problem with relearning, as suggested by Martin and Deutscher, is that the source of the person’s mental states about the event is not the event itself, but some later event during which forgotten information was re-acquired. Given that the information is accurate, by Martin and Deutscher’s design, there is one sense in which relearning is difficult to characterize as an error (for this reason Bernecker 2017 has argued for removing it from the taxonomy of memory errors). But instances of relearning are errors insofar as they are treated as memories rather than second-hand reports. They are often a form of what cognitive psychologists would call source-monitoring errors.

In what follows, I lay out my version of the Causal theory and Michaelian’s (2016a; 2016b) version of the Simulation theory. Each includes three conditions and illustrates the ways in which meeting or failing various combinations of them yields distinctive memory errors. After presenting each view I lay out two cases, which I stipulate as instances of remembering and confabulation, respectively, and argue that the Simulation theory misclassifies each of them. These failures illustrate the limitations that come from an appeal to reliability in an account of remembering (or so I will argue).

3.2 Robins Causal Theory

Developing an account of episodic remembering requires evaluating the set of mental states that involve seeming to remember, sorting the cases of successful remembering from memory error. Seeming to remember, as I define it here, occurs when a person has an occurrent mental representation, the content of which targets a representation in her personal past. Seeming to remember episodically

---

6 For example, the comparison to illusion and hallucination can be found in both Robins (2016) and Michaelian (2016b), even though we differ in our characterizations of both misremembering and confabulation.

7 The arguments, and much of the text here, are taken from Robins (2019).
picks out a set of mental representations that share a target type, which I term a **3P event**: the target is a *particular* event or experience in the representer’s *personal past*.

To determine whether any particular case of seeming to remember qualifies as an instance of successful remembering requires a more fine-grained evaluation of the target and its relation to the content of the occurrent representation. Let’s consider an example. Jamari is in a state of seeming to remember—he has a mental representation of opening a gift at his college graduation party; inside the box is a beautiful watch. This counts as an attempt at remembering because he is targeting a particular event in his personal past. The target is the graduation party. The content is the information contained in the occurrent mental representation he forms. Jamari’s representation would likely include details about the gift—that it was a small box, wrapped in blue and silver paper, containing a watch with a dark leather strap. These contents might be represented as propositions or as images. The representation might also include physical sensations, like the feel of the wrapping paper and the weight of the watch, and other phenomenological features or emotional qualities.

There are three conditions that must be met for this, or any, instance of seeming to remember to be actual remembering:

1) **Target.**
   The first condition is a stipulation about the targeted event: it must exist. Jamari must have had a graduation party. The existence of the target is critical not only because memory is factive, but because the targeted event frames the evaluation of the next two conditions on remembering.

2) **Accuracy.**
   The content of the mental representation must be accurate. Accuracy is not simply a matter of the representation being true, but of its being true of the event targeted. Jamari

---

8 The event targeted here could be specified more finely: perhaps it’s the opening of a gift at this party rather than the entire party. In further development of this view, I will have to say more about how targets are individuated—and what I have to say will be informed by empirical work on event processing (most especially Event Segmentation Theory, Kurby & Zacks, 2008). For now, I think that it is enough to say that the scope of an event can, and likely will, vary across contexts. Getting the target right is important, of course, because how the target is understood will influence our determinations of when a memory is in error. Thanks to an anonymous reviewer for encouraging me to make this feature of the account more explicit.

9 I avoid endorsing any particular view about the nature of mental content in general or the contents of episodic memory in particular. My aim is to remain ecumenical, as what matters for my purposes is the relation of the target and the content, however that content is understood.

10 Much more needs to be said about how to determine the accuracy of memory. Bernecker (2010) makes a helpful distinction between truth and authenticity, where truth refers to the objective veridicality of the memory state and authenticity refers to its relation to the person’s earlier
may have received such a watch, but if he received it for a birthday rather than graduation, it would fail to be an instance of successful remembering.

3) **Causal History.** The third requirement is distinctive of the causal theory: the content of the representation must have been produced in the right way, where the ‘right way’ involves a causal connection between the original event and the current representation, and more specifically, a causal connection maintained by a memory trace (a mental and/or neural mechanism for retaining memories). Often causal theorists articulate this requirement as the need for an unbroken causal path between the event and the subsequent remembering. I prefer instead to think of the condition as a requirement on the causal history of the mental process by which the representation is produced. Jamari’s representation of the watch now must have been produced by a memory trace he formed as a result of the graduation party.

Memory errors occur when one or more of these conditions are violated. I discuss three memory errors below: misremembering, relearning, and confabulation. These errors can involve violations of multiple conditions, but each error is characterized most essentially by a failure to meet one of the three conditions. Table 1 displays the requirements on remembering, illustrating how they are involved in successful remembering, misremembering, relearning, and confabulation.

**Misremembering**

Misremembering is an error that concerns the accuracy of a memory’s content. It arises because of a mismatch between the content and the target. Robins (2016) misremembering as follows:

```
Misremembering is a memory error that relies on successful retention of the targeted event. When a person misremembers, her report is inaccurate and yet the error is explicable only on the assumption that she has retained information from the event her representation mischaracterizes (p. 434).
```

This form of misremembering can be illustrated by tweaking the Jamari example from above. Suppose Jamari is in a state of seeming to remember his college graduate party. His representation targets that event in his personal past, but the content of his representation is not entirely accurate. Maybe his representation depicts the wrapping paper on the gift box as gold, rather than blue and silver, or depicts the gift as a fountain pen rather than a watch. Any of these would be errors, and would keep this from being an instance of successful remembering. Still, the representations are

representation of the event. Bernecker claims that both are required for veridicality. An alternative, made possible by the framework I have introduced here, would be to allow that features of how the event is being targeted (e.g., the event as it occurred or the event as it was experienced) can be used to assess the accuracy of the representation.
mostly accurate; the inaccuracies involved are distortions of the actual event. This form of misremembering is illustrated most clearly in the errors generated by the DRM paradigm in cognitive psychology (Roediger & McDermott 1995; see also Robins 2016).

There may also be other forms of misremembering, which stretch the original definition above. Misremembering could occur when there is a mismatch between the target and the content, as when the target of Jamari’s representation is the graduation party, but the content is drawn from the opening of a gift at a different celebration, like a birthday. In such a case, accuracy is still the primary concern. The content generated is an accurate representation of some event in Jamari’s past, but not the event he takes himself to be representing and remembering.

**Relearning**
Relearning errors occur when there is a problem with the causal relationship between the content generated and the event being targeted. In cases of relearning, a person has an experience, forgets it, and then learns of it from somewhere else, and this later relearning is at some point confused for remembering. Suppose Jamari received a watch at his college graduate party, but forgot all about it, either through natural means or because of some kind of neurological trauma that produced amnesia. He later discovers a video of the party, including his opening of the watch. Over time, he forgets how he acquired this information about receiving a watch at his graduation and when he thinks about this party, he takes the mental representations he forms to be a memory of the party (rather than a memory of watching a video of the party). In such a case, the target of Jamari’s representation is the graduation party, and the content of his mental representation accurately depicts that event. The error is in the relationship between the target and the content. It seems to Jamari that he remembers his graduation party, but he does not. Relearning can be understood as a form of memorial hearsay, where information about an event or experience is misinterpreted as being from an experience.

**Confabulation**

---

11 In standard cases of relearning, as originally introduced by Martin and Deutscher (1966), what is relearned is true or accurate of the event. The case is meant to illustrate the importance of causal history for remembering, and so relearning is meant to be a case where all other requirements (including accuracy) are met. Michaelian (2016b) has argued that taxonomies of memory errors should also accommodate falsidical relearning. Although I have not built it in explicitly, falsidical relearning is consistent with the account of remembering and its errors that I am developing here.
Confabulation occurs when there is no relation between a person’s feeling as if they remember a particular event/experience and any event or experience from their past—either because there is no such event in their past or because any correspondence to such an event is entirely coincidental. Where cases of relearning and misremembering involve some form of mismatch between the target and content, cases of confabulation involve no connection at all. In this way, confabulation is a memory error that is concerned most directly with the requirement that the targeted event must exist. In most cases of confabulation, it does not. Suppose Jamari never went to college, and has in fact spent most of his adult years living in a psychiatric facility. He could enter a state of seeming to remember receiving a watch at his college graduation party. This instance of seeming to remember would be a confabulation because there is no event in his personal past corresponding to the one he is targeting. This is the form of confabulation that has generated the most clinical and theoretical attention and interest. Many definitions of confabulation have even made the falsity of the memory an essential feature of confabulation. But, as others have noted, confabulations could be veridical. In these cases, the targeted event exists, but there is no connection between the target and the content. To see the point, we can return to the form of the example where Jamari graduates from college, and has a party afterward where he receives a watch. He then forgets about this experience, perhaps as the result of some amnesia-producing trauma. At some later point in life, saddened by the memories he has lost to amnesia, he begins inventing stories about his past. After frequent retelling, he begins to consider many of these stories genuine memories. One of these stories, as it happens, matches exactly his experience of receiving a watch at his college graduation party. The representation is an accurate depiction of an event from his personal past, but this does not come about because of any information Jamari has about the graduation party (either from his memory or a video), as Jamari has lost this information and has not reencountered it in any other context. There is no connection between the event and his representation of it. It is just an instance of serendipitous confabulation.

3.3 Michaelian’s Simulation Theory

Michaelian’s (2016a; 2016b) simulation theory is one of the more prominent of a set of recent philosophical views of remembering that have developed in attempt to account for the broad trend in memory science to think of remembering as constructive (see also De Brigard 2014; Hopkins 2018). Constructive views about the nature of memory have come about largely in response to the preponderance of evidence about memory errors in cognitive psychology and neuroscience.
Accumulating evidence that false memories are commonplace in everyday remembering has led many to urge a rethinking the nature and function of memory. Constructivists argue that memory is a capacity for building (i.e., constructing) representations of past events from a generalized network of information. Precisely which features of this general characterization are emphasized, and how, differs across the view’s proponents.

Constructivists argue that the accumulation of such errors presents a serious challenge to the traditional view of memory, placing pressure on its preservative portrayal. It is strange to characterize memory as a capacity designed for preservation if it rarely manages to keep memories of past events intact. As De Brigard explains, “saying that false and distorted memories are a failure of memory may force us to accept that we have a memory system that regularly and systematically malfunctions” (2014: 159). A better option, the Constructivist claims, is to think that we have misunderstood memory’s function. If we instead view memory errors as the primary explanandum—asking what purpose these false and distorted representations of the past might serve—then what emerges is an altogether different account.

So what is this alternative account of memory’s function? The characterization is usually contrastive, selecting a feature of the traditional preservationist picture against which the Constructivist alternative can be understood. The general Constructivist approach is to situate remembering within a broader cognitive system, as one use of an episodic (i.e. self-directed) capacity amongst many. De Brigard’s (2014) *Episodic Hypothetical Thinking* advocates for a change in the goal of memory processing, to *plausibility* rather than *accuracy*. This change in function fits memory’s place in a larger cognitive system devoted to episodic hypothetical thinking—a system for “self-referential mental simulations about what happened, may happen, and could have happened to one self” (2014: 174–175). Michaelian’s (2012; 2016a) *Simulation Theory* presents a similar but distinct version of Constructivism. There are two key differences between De Brigard and Michaelian’s versions of Constructivism. First, they offer distinct portrayals of the constructive process. Michaelian characterizes the broader cognitive system as one dedicated to producing episodic imagination—i.e., the simulation of possible episodes. The only difference between remembering and other forms of episodic imagination is the aim of the simulation process. In remembering, the aim is to produce an episode from the rememberer’s past (2016: 105). Second, they differ over whether the constructive system is characterized as aiming at and maintaining accuracy in its constructions. While De Brigard says no, Michaelian says yes. These episodic simulations, whether intended for imagining a possible
future or remembering the personal past, will function at their best when the cognitive system is streamlined. And so, Michaelian’s view helps to make sense of why forgetting occurs and why it may be best construed as an epistemic virtue:

Given limited computational resources, forgetting is necessary to enable the system to achieve the balance of reliability, power, and speed appropriate for it given its function; given the necessity of sensitivity to interests for virtue, virtuous memory involves preferentially forgetting uninteresting records (Michaelian, 2011: 423).

The Simulation Theory also makes sense of why memory errors occur. Michaelian argues that most of the additional content utilized in the construction of episodic remembering serves as information rather than misinformation. Nonetheless, a system that aims at simulating episodes from one’s personal past may occasional fail to do so; incorporating misinformation into one’s simulation can lead to error.

Michaelian’s (2016) Simulation Theory offers a different approach, one that allows for a distinction between memory errors consistent with the episodic system’s functioning and memory errors indicative of malfunction. Michaelian draws the distinction in terms of the system’s tendency toward accuracy. As he explains:

In a healthy subject, the system recombines information, whether or not it originates in the target episode, following procedures designed to enable it to produce a representation of the episode which is (within certain limits) accurate. In the confabulating subject, in contrast, the system malfunctions, following procedures which tend to produce inaccurate representations (2016: 109).

Michaelian goes on to explain that memory’s tendency toward accuracy is sustained in two ways. First, the simulation process is governed by heuristics that favor information over misinformation. Second, the simulation process is overseen by a source monitoring system whose job is to reject simulations that have gone too far afield. These mechanisms ensure that the system is by and large reliable, but the mechanisms are not infallible. Their operation is consistent with the occasional memory error. Michaelian’s approach thus distinguishes everyday memory errors from confabulations by the differences in reliability for the cognitive systems in which they occur.

In his 2016b, Michaelian portrays his account of remembering in terms of three conditions used to separate remembering from the errors of confabulation, misremembering, and relearning. To do so, he appeals to three conditions: accuracy, internality, and reliability.
The accuracy condition does not get much in the way of elaboration, in part because of the ways in which it is both intuitive and shared between Simulationist and Causalist. There may be differences in how much accuracy is required or how it is determined, but the general expectation is that the condition is the same.

The internality condition is a way of acknowledging the fact that memory is a capacity for retention. In order to count as remembering something a person must contribute, via information they have retained, to the representation of information about a particular past event. It’s not unusual for instances of remembering to involve some degree of prompting—maybe you need a bit of a nudge from a friend or from your notes in order to call the event to mind. Needing help to “jog one’s memory” is commonplace and allowable, on both Causalist and Simulationist views.\textsuperscript{12} The concern with the internality condition is for cases where the information is entirely external and not within or due to the agent in any sense. Michaelian does not define internality, other than to characterize cases of its failure as ones where the subject “contributes no content to the retrieved representation” (2016b: p. 5). Here Michaelian is thinking about cases of relearning as instances where the internality condition is violated. When the subject has to rely on outside sources in order to reacquire information about a past experience, it fails as an instance of remembering because the information is not internal.

The reliability condition is meant to give us a way to sort between cases of misremembering and confabulation. Failures of the accuracy condition are not enough to characterize the distinction between them, as both are inaccurate (at least most of the time – although veridical confabulation is possible), just to differing degrees. For the Causalist, the difference in the amount of error is explained by appeal to the information retained from the particular past event and how much of a role it plays in the explanation. Since the Simulationist does not believe that information from the past is available in this way, he or she must look for an alternative. This is where the notion of reliability comes in. Michaelian notes, explicitly, that he is making use of the epistemologist’s notion of reliability, as he says “Reliability is understood here in the epistemologist’s sense: a reliable system is, roughly, one that produces mostly accurate representations” (2016b: 6). Michaelian goes on to make clear that he intends a modal rather than frequentist interpretation of the systems’ reliability (although not to give an expression of that modal claim).

\textsuperscript{12} There are differences in how cues and prompts are treated on distinct Causalist views – e.g., Martin and Deutscher (1966), Bernecker (2010) and Robins (2017) each offer different proposals.
Michaelian considers it a virtue of his view that he can distinguish not only the errors of misremembering, confabulation, and relearning, but that his account can sort between veridical and falsidical versions of each.

### 3.4 Evaluating Tough Cases

My version of the Causalist view and Michaelian’s Simulationism, and how they account for and explain various memory errors can be seen in the visual displays at the end of the paper (Table 1 and Figure 1, respectively). Each account involves three conditions. Although they do not align perfectly, it’s clear to see that the Causal History condition in my account and the Reliability condition in Michaelian’s play equivalent roles in the respective accounts.

In this section, I introduce a few hypothetical but plausible cases that I will use to evaluate the two accounts of remembering and error outlined in the previous sections. In doing so, I’m assuming broad agreement on how these cases should be interpreted – more specifically, that the first case will strike most if not all of us as a case of remembering and that the second case will strike us as an instance of memory error (confabulation). Michaelian’s view, I will show, does not classify either of these cases in this way, thus violating the intuitive interpretation of each.

**Case 1: The Dementia Patient**

Consider the case of Claire, a woman in her late 70s with Alzheimer’s Dementia. Claire was diagnosed many years ago and her condition is progressing from the middle stage to the late (and final) one, from moderate to severe dementia. At this point, she is sometimes confused about basic facts of her daily life – where she is, what day it is, who is President, etc. She also routinely forgets details of her personal history and is becoming less responsive to reminders and prompts from those around her. Still, she enjoys telling stories of her past and has one story in particular that she recounts somewhat regularly—the tale of winning an essay prize in her final year of undergraduate study. She recalls how hard she worked on the essay in the weeks leading up to the deadline, how nervous she felt and how confident she was that the award would go instead to James who had bested her in various other academic competitions throughout college. Her story involves, too, the warm feeling of the lights from up on the stage and the moment after her name was called where she stood still and confused, not entirely believing what she’d just heard. She recalls hugging some faculty members and celebrating with friends after. She can even say a few things about the content of the essay and why she’d been so thrilled to write it.
Does Claire remember winning the essay prize? We'll assume that she did in fact win such a prize, meeting the conditions of accuracy and targeting/internality. For the Causal theory, the assessment of whether Claire remembers is going to depend on the kind of connection her representation of this event bears to the experience of the event itself. So long as there is such a connection, then yes, she is remembering. For the Simulationist, the question is whether her account derives from a reliable memory system. As someone in the late stages of dementia, Claire’s memory system is anything but reliable. And so, following Fig. 2, this case would be classified as one of Veridical Confabulation.

I take it that Claire is remembering in this case, or at least if we could have a metaphysics-eye-view of the scenario, we’d find evidence to this effect. At least, it’s consistent with having an overall malfunctioning or improperly working system, that there are still some instances of success. Maybe many of those who interact with Claire daily (nurses, doctors, therapists, etc) will be skeptical of this memory’s legitimacy and treat it as they do the other stories she generates. They would have good reason to do so, given that she hasn’t been particularly reliable in remembering as a function of her illness. Nonetheless, it is still possible for her to remember – and being present when she tells such a story could give hope or comfort to a family member or friend who comes to visit, as evidence that at least some memories are “still in there.”

One might want to raise challenges to this case, in particular, the reliance on a case of disease or disorder for making claims about how systems are working. But this is the kind of scenario we will have to establish in order to have a system that is, overall, unreliable. The point is that the general unreliability of the system is no good reason for saying that every particular instance will fail to be an instance of remembering. Especially in progressive disorders such as dementia, the gradual shift in proportion is the kind of thing we might like to account for.

*Case 2: Highly-Superior Autobiographical Memory*

Sam is a man in his early 40s who has been diagnosed with Highly-Superior Autobiographical Memory. For most of the days of his life since early adolescence, Sam has a vivid and accurate memory of what transpired on that day. People with HSAM are somewhat rare, but not entirely uncommon (McGaugh et al., 2006; Mazzoni et al., 2019).

One of Sam’s memories includes helping a person change their flat tire on the side of the road, on his way to work one morning. Sam recounts the make and model of the car, the person’s distress, and the conversation they had while trying to remove the lug nuts. All of the details of this
story look comparable to the details of other stories that Sam has told (and have been verified by the researchers who study him), but in this case, Sam’s memory is inaccurate. No such event occurred.

According to the Causalist view I’ve outlined, Sam is confabulating. There is no connection between this event and any event in his past. In other words, the event Sam is targeting does not exist. For Michaelian, such a case would qualify as an instance of misremembering. The content is inaccurate, but it’s the result of a memory system that is, overall, highly reliable.

Generalizing from this case, it looks like a problem for Michaelian’s view that confabulations can only occur for those who have unreliable memory systems. This is a problem because the possibility of veridical confabulation – of coincidentally landing in a situation where one memorially hallucinates an event just like one that was previously experienced – is possible for each of us (even if highly unlikely). Nothing in Michaelian’s account allows for this possibility.

These two cases were designed to put the overall reliability of the memory system at odds with the performance in a particular case. I did so in order to put pressure on Michaelian’s attempt to collapse these two judgments in an account of remembering by using the overall reliability of the system as a way to gauge whether or not remembering has occurred. If I am correct about how we want to classify and judge such cases, then I take it I have shown why reliability cannot do the necessary work. It cannot replace a memory trace requirement in an account of remembering and capture the judgments we would like to make on the full range of possible cases. These failures, I want to go on to suggest, do not derive from any particular faults or limitations in Michaelian’s argument – they’re instead direct consequences of the appeal to reliability, which is not designed for this work.

4. Two Lessons/Conclusions

In this final section I want to make two points, which are a way of combining the distinctions made in Section 2 with the argument against Michaelian’s Simulationism in Section 3. That is, I want to use the points made about the differences between epistemological issues with memory and accounts of remembering and the failures of Michaelian’s reliability-based account of Simulation to make some broader points about why reliability cannot do the work that Michaelian has set out for it.
1. Assessments of Reliability are Statistical; Assessments of Remembering are Particular

Reliability is a statistical notion. It is a judgment or determination about a process or system or capacity that is made on the basis of a set of observations of its behavior or performance (actual past behavior, possible behavior across worlds, etc.). It involves, critically, judging the system’s performance in a particular instance on the basis of how it has or would behave across a broader range of instances. For judgments about a number of cognitive capacities and abilities, this is quite useful. It is not, however, helpful when assessing remembering. This is because assessments of remembering are particular. The question is about the relationship of this particular mental state to some particular event or experience in my past. Of course we can make judgements about the overall reliability of my memory system. And if my system is unreliable, then you might have good reason to bet against my remembering in a particular case. But there is no influence from the general system to the individual state. The assessment of the system’s reliability is, at best, additive.

Previously, I’ve been inclined to make this argument against particular forms that reliabilism could take. In Robins (2018) I argued against a frequentist understanding of Michaelian’s appeal to reliability, based on the view put forward in Michaelian 2016a. There I argued that the frequency notion wasn’t well suited to analyses of remembering, given that one’s remembering that p often occurs only once. I was also concerned that we lacked clear evidence that those with serious memory malfunctions (like confabulation due to a broader psychiatric disorder) actually produced more memory errors than those of us without such a memory deficit. I urged instead a view that distinguished between types of errors, rather than frequency or amount of error.

Since I wrote this paper, Michaelian has gone on to clarify that the notion of reliability he has in mind is a modal rather than frequentist one (2016b). Similar concerns can be raised for this characterization of reliability, too. It is unclear to me how we would determine the set of relevant possible worlds for evaluating S remembers p. The point of departure from general epistemic practice concerning reliability matters here, as the assessments made over relevant worlds won’t be ones of whether p is true in those worlds, but whether it’s true that S remembers p in those worlds. Any way we vary the case from S remembers p to some similar but distinct variant in a possible world changes the case in ways that make it seem less relevant for assessing the original claim. Suppose we maintain everything but change the time (relative to S’s lifespan) when the representing p occurs. Regardless of whether S turns out to be remembering here or not, it’s unclear how it bears on the mental state in the actual world. The success of my remembering something now is not downgraded by my
failure to remember it a few hours ago (although my exam grade might be). So too for assessments of my remembering $q$ or even other variants of $p$. The case only gets worse if we move toward worlds that are even more dissimilar.

I’m no longer inclined to make the argument via the details of particular reliabilist programs. The general point that unites all of these arguments is the simpler and more straightforward one about the scope of assessment. Reliability is a claim about a system as a whole; Remembering is a claim about a particular mental state or act.

2. **Belief cross-contaminates; Remembering does not**

Some of the motivation for reliabilist moves in epistemology comes from the felt need to add a more global or holistic perspective to assessments of true beliefs. The worry I have in mind is one about lucky beliefs or beliefs that turn out to be true in some roundabout and bizarre way such that there is a strong resistance to labeling them as instances of knowledge. Here the general thought seems to be that the belief’s status with regards to knowledge is determined in part by the overall system or by the other beliefs with which it associates. We care about *how* the belief was produced because we want attributions of knowledge to be deep and systematic and thus more likely to hold fast when circumstances change and to iterate in the further beliefs formed in similar ways. Even if two beliefs are equally true, differences in their production have an influence on our assessments of their value.

In this way, our assessment of the value of belief is quite different than our assessment of the value of other systems in the world. To borrow Zagzebski’s way of making the point (as many do), this isn’t how we treat our coffee machines. A good cup of coffee (or espresso) is equally good to us regardless of whether it came from a generally faulty and unreliable machine or from a highly reliable one. The value is in the coffee itself, not in the overall system from which it came. Of course things are different when we’re shopping for a new coffee maker, but the way in which we care about the system here is simply an additive extension of the point already made. We want multiple good cups of coffee so we want the system that will give us the most such cups. There is no over-and-above value to being a coffee machine built to do this in a reliable and systematic way.

Thinking through the similarities and differences here is really helpful to me in thinking through what one wants in developing an account of remembering. There are some ways in which memory is like belief here, in that we care about the nature of the process by which it was formed. There is value in that over and above the truth or accuracy of the memory. But what’s valued is not
the memory system generally, as is the case for belief, but the particulars of how the memory connects us to a particular past event or experience. Memory is treasured for the connection it gives us to our past, which is thought to be something more or more significant than having true beliefs about our past experiences. Perhaps Claire, the Alzheimer’s patient in an above example, can be retaught about many of her past experiences. We might even suppose that she was an extensive diarist and we can use those notes, along with today’s more extensive technology, to give her vivid details and imagery to go along with her beliefs. This is likely to strike many of us as better than nothing, but still something short of remembering.

Taking this point further, there are other ways in which remembering is much more like a cup of coffee. The ability of an instance of remembering to be such an instance is not influenced by the general system from which it comes. Many of my true beliefs about the past may be ones that I come at piecemeal—a combination of inference, relearning, and fused generalizations across multiple experiences. Still, a few others connect me back to particular past experiences. The proportions of each likely differ amongst us, and differ for any of us at distinct stages of our lives. Maybe this brings about changes in how much we trust our memory’s allegations about the past, but it doesn’t change the status of those states as states of remembering. Memory’s value is in its particularity.

The Simulationist wants to resist the need for particular connections to past experiences (via memory traces) as a requirement for remembering, but they do want to carve out a unique set of beliefs and claims about the past that are remembering. Not every accurate mental state about one’s past experiences counts as remembering, at least not so far as I can tell. So it’s possible that Michaelian and others will want to push back on the full-strength version of my claims here and carve out a different space for remembering. To do so they will have to be more careful and creative in the ways that they pull over the notion of reliability, as it is too general and systems-level to make the needed distinctions.

In making the points I’ve made in this paper, it’s hard to resist the feeling that I am simply begging the question against the Simulationist. The aim of Simulationism, and other constructive views of memory, is intended as revolutionary and it’s possible that my arguments only serve to illustrate just how much they seek to upend. Maybe that’s so. But even if it is, there is (I hope) some good in making explicit just how many claims and assumptions about remembering the Simulationist is rejecting alongside the memory trace.
References

### Table 1. Robins Causalist Taxonomy

<table>
<thead>
<tr>
<th></th>
<th>1) Target Exists</th>
<th>2) Accurate Content (content = target)</th>
<th>3) Causal History (content produced by target memory trace)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Remembering</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Misremembering</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Relearning</td>
<td>Yes</td>
<td>Yes</td>
<td>No, causal history is deviant</td>
</tr>
<tr>
<td>Confabulation</td>
<td>Yes or No</td>
<td>Yes or No</td>
<td>No, there is no causal history</td>
</tr>
</tbody>
</table>

### Figure 1. Michaelian’s Simulationist Taxonomy (2016b)

![Simulationist Taxonomy Diagram]

*FIGURE 1 | The simulationist taxonomy of memory errors.*