



# Imagining the past reliably and unreliably: towards a virtue theory of memory

Kourken Michaelian<sup>1</sup>

Received: 14 June 2020 / Accepted: 13 March 2021 / Published online: 3 April 2021  
© The Author(s), under exclusive licence to Springer Nature B.V. 2021

## Abstract

Philosophers of memory have approached the relationship between memory and imagination from two very different perspectives. Advocates of the causal theory of memory, on the one hand, have motivated their preferred theory by appealing to the intuitive contrast between successfully remembering an event and merely imagining it. Advocates of the simulation theory, on the other hand, have motivated their preferred theory by appealing to empirical evidence for important similarities between remembering the past and imagining the future. Recently, causalists have argued that simulationism is unable to accommodate the difference between successful remembering and forms of unsuccessful remembering or mere imagining such as confabulating. This paper argues that, while these arguments fail, simulationism, in its current form, is indeed unable to provide a fully adequate account of unsuccessful remembering. Rather than suggesting a return to causalism, the paper proposes a new form of simulationism, a virtue theory of memory modelled not on the process reliabilist epistemology that has so far served as the inspiration for the simulation theory but instead on virtue reliabilist epistemology, and shows that this new theory grounds a more adequate account of unsuccessful remembering.

**Keywords** Memory · Imagination · Causal theory of memory · Simulation theory of memory · Confabulation

*It is possible to fail in many ways, while to succeed is possible only in one way.*  
-Aristotle (trans. W. D. Ross)

---

✉ Kourken Michaelian  
kourken.michaelian@univ-grenoble-alpes.fr

<sup>1</sup> Centre for Philosophy of Memory, Institut de Philosophie de Grenoble, Université Grenoble Alpes, Bâtiment ARSH, 1281 Avenue Centrale, CS 40700, 38058 Grenoble Cedex 9, France

## 1 Introduction: memory and (two kinds of) imagination

The relationship of *episodic memory* to *episodic imagination* is among the core concerns of the philosophy of memory (De Brigard, 2017; Liao & Gendler, 2019). When philosophers of memory refer to “imagination”, they typically assign one of two senses to the term, both meant to contrast imagination—at least potentially—with memory. “Imagination” can mean, first, *imagination-as-failure*: one can attempt to remember an event and seem to do so but nevertheless fail to do so; in such cases, one is sometimes said merely to imagine the event that one apparently remembers. Memory and imagination, in this first sense, are distinct potential outcomes of the retrieval process. “Imagination” can mean, second, *imagination-as-process*: one can (successfully) remember an event, but one can also (successfully) imagine it; on the face of it, at least, it is one thing to remember an event and another to imagine it. Remembering and imagining, in this second sense, seem to be distinct cognitive processes.

Both of these senses of “imagination” figure in the ongoing controversy between causalist and simulationist philosophers of memory. The *causal theory of memory* (CTM) is motivated primarily by its apparent ability to distinguish between remembering an event and imagining it, in the imagination-as-failure sense (Martin & Deutscher, 1966; see also Bernecker, 2008, 2010; Debus, 2017). The *simulation theory of memory* (STM), CTM’s main rival, is motivated primarily by empirical evidence of important similarities between remembering the past and imagining the future, evidence that simulationists take to suggest that memory is a kind of imagination, in the imagination-as-process sense (Michaelian, 2016c; cf. Shanton & Goldman, 2010; De Brigard, 2014a).<sup>1</sup> There are ongoing debates between causal theorists and simulation theorists over both of these issues. Causalist responses to simulationist claims regarding the relationship between remembering the past and imagining the future have led to the *continuism-discontinuism debate*, in which causalists have contested simulationists’ interpretation of the empirical evidence that they take to support STM, pointed to apparently contradictory empirical evidence, or argued for metaphysical or epistemic differences between remembering the past and imagining the future, while simulationists have defended STM’s empirical credentials and questioned the existence or downplayed the importance of the relevant metaphysical and epistemic differences (Perrin, 2016; Michaelian, 2016a; Perrin & Michaelian, 2017; Michaelian, Perrin & Sant’Anna 2020; Robins, 2020a; Munro online ahead of print). Simulationist responses to causalist claims regarding

<sup>1</sup> Throughout the paper, “CTM” and “STM” refer to the specific causal and simulation theories stated, respectively, in Sects. 2.1 and 2.2. “Causalism” and “simulationism” refer to broader approaches both to genuine remembering (the target of CTM and STM) and to merely apparent remembering. Causalism and simulationism are thus broader than CTM and STM in two senses. First, multiple causalist and multiple simulationist classifications of memory errors are available. The main available classifications are discussed below. Second, multiple causal theories and multiple simulation theories are available. Regarding the availability of different causal theories, see Michaelian and Robins (2018). Regarding the availability of different simulation theories, see the discussion of the relationship between VTM and simulationism below.

the relationship between successful remembering and unsuccessful remembering or mere imagining, meanwhile, have triggered the *memory error debate*, in which simulationists and causalists have defended rival analyses of confabulating and related forms of unsuccessful remembering (Baysan, 2018; Bernecker, 2017; Michaelian, 2016b, 2020; Robins, 2016a, 2019, 2020b).

This paper intervenes in the latter debate. Building on a review of rival causalist and simulationist approaches to unsuccessful remembering, it argues that, while causalist arguments for the claim that simulationism is unable to accommodate the difference between successful and unsuccessful remembering fail, simulationism, in its current form, is indeed unable to provide a fully adequate account of unsuccessful remembering. Rather than suggesting a return to causalism, the paper proposes a new form of simulationism, a *virtue theory of memory* modelled not on the process reliabilist epistemology that has so far served as the inspiration for the simulation theory but instead on virtue reliabilist epistemology, and shows that this new theory grounds a more adequate account of unsuccessful remembering.

The paper is structured as follows. Section 2 reviews the simulationist attack on CTM on the basis of evidence of similarities between remembering the past and imagining the future. Section 3 discusses the causalist attack on STM on the basis of its apparent inability to provide an adequate account of confabulation, arguing that the account of confabulation suggested by simulationism is in fact preferable to the account suggested by causalism. Section 4 argues, however, that that account implies that STM itself is inadequate and argues for the replacement of STM by a virtue theory of memory. Section 5 concludes with a brief discussion of the possibility of a broader virtue theory of imagination.

## 2 Remembering as reliable imagining

This section provides a brief review of the simulationist attack on CTM on the basis of evidence of similarities between remembering the past and imagining the future. The following section turns to the causalist attack on STM on the basis of difficulties apparently faced by simulationism in accounting for the relationship between successful and unsuccessful remembering.

### 2.1 The causal theory of memory

Martin and Deutscher's (1966) influential statement of the causal theory<sup>2</sup> employs some now-unfamiliar vocabulary, but the theory can be given a straightforward formulation:

---

<sup>2</sup> Recent neoclassical causal theories, including those defended by Bernecker (2008, 2010) and Cheng and Werning (2016), differ from Martin and Deutscher's classical account in interesting and important ways, but these differences—on which see Michaelian and Robins (2018)—will not matter here. Werning's (2020) more recent "minimal trace theory" is, however, discussed in Sect. 3 below.

- (CTM) A subject S now remembers an event e iff  
 (PE) S experienced e when it occurred;  
 (CR) S now represents e;  
 (AC) S's current representation of e is appropriately causally connected to S's previous experience of e, where an appropriate causal connection is one that is sustained by a memory trace originating in S's experience of e.

The *previous experience* and *current representation* conditions (PE and CR) are relatively uncontroversial.<sup>3</sup> The distinctive feature of CTM is thus its inclusion of the *appropriate causation* condition (AC). The basic strategy of Martin and Deutscher's argument for CTM is equally straightforward: they argue, appealing to our intuitions about a series of hypothetical cases, for the necessity of AC; they then argue, again appealing to our intuitions, for the joint sufficiency of PE, CR, and AC.

In order to establish the *necessity* of AC, Martin and Deutscher ask us to consider a case in which a subject experiences an event, loses all memory of it, and then comes, under the influence of a hypnotist with no knowledge of the event, to entertain a representation that happens to be accurate with respect to the event in question. They next ask us to consider a case in which a subject experiences an event, recounts it to a friend, loses all memory of it, is told about the event by the friend to whom he recounted it, loses all memory of being told, and then comes, under the influence of what he has been told, to entertain a representation that happens to be accurate with respect to the event in question. Intuitively, the subject remembers in neither case. The unusual feature of the hypnotist case is that there is no causal connection between the subject's current representation of the event and his previous experience of it. We are thus invited to draw the conclusion that it is necessary for remembering that there be a causal connection between the subject's current representation and his previous experience. The unusual feature of the friend case is that, though there is a causal connection between the subject's current representation of the event and his previous experience of it, the connection is not sustained by a memory trace originating in the experience. We are thus invited to draw the conclusion that it is necessary for remembering not just that there be some causal connection or other between the subject's current representation and his previous experience but that there be a causal connection sustained by a memory trace originating in the experience.

In order to establish the joint *sufficiency* of PE, CR, and AC, Martin and Deutscher ask us to consider a case in which a subject paints a scene that he takes to be imaginary but that exactly resembles a scene that he experienced as a child. Intuitively, if there is an appropriate causal connection between the subject's current representation of the scene and his previous experience of it, he is remembering. We

<sup>3</sup> Note, however, that, while simulationists accept CR, they may reject PE, reasoning that, if remembering is a kind of imagining, there is no reason to suppose that one can remember only events that one previously experienced. As formulated by Michaelian (2016c), STM thus implies that one can potentially remember an event even if one did not experience it. Somewhat surprisingly, causalists have not so far argued in any detail against STM on the basis of this counterintuitive implication.

are thus invited to draw the conclusion that PE, CR, and AC are jointly sufficient for remembering.

## 2.2 The simulation theory of memory

CTM long occupied a dominant position in the philosophy of memory, but the theory has been subjected to an increasing number of challenges. Some of these challenges, while important, do not get at the heart of the theory, namely, the claim that remembering requires appropriate causation. There have, for example, been challenges (based on the research on the constructive character of remembering discussed below) to *preservationism*, the view that remembering requires that the retrieved apparent memory not include any content not included in the corresponding earlier experience (Michaelian, 2011).<sup>4</sup> But while Martin and Deutscher seem to have been preservationists, and while most contemporary causalists continue to endorse preservationism (e.g., Bernecker, 2008, 2010; McCarroll, 2018), causalism does not—since a causal connection between a retrieved apparent memory and an earlier experience might be sustained by a memory trace even if the apparent memory includes content not included in the experience—entail preservationism. These challenges thus do not get at the heart of the theory.<sup>5</sup> There have also been challenges both to CTM’s reference to *memory traces* (Malcolm, 1977) and to the specific conception of memory traces (as structural analogues of experience) with which Martin and Deutscher worked (Sutton, 1998; Robins, 2016b, 2017a, b; Perin, 2018; see De Brigard, 2014b, 2020). But other theories, including STM, likewise refer to traces: simulationists reject the strong claim that remembering requires appropriate causation but, like causalists, invoke traces in their explanation of the workings of memory.<sup>6</sup> These challenges, too, thus do not get at the heart of the causal theory.

Other challenges, however, focus squarely on appropriate causation. There have been challenges based on the *epistemic theory of memory*, according to which

<sup>4</sup> Preservationism is sometimes defined as the view that memory is preservative in two senses: first, the content of a retrieved memory may not exceed the content of the corresponding earlier experience; second, remembering is not constructive. McCarroll, for example, defines preservationism as “the idea that memory preserves perceptual content *and* stores static items for later retrieval” (2018, p. 12; emphasis added). As Michaelian & Sant’Anna (manuscript) point out, however, there are two distinct ideas here. The first idea might entail the second: the content of a retrieved memory presumably cannot exceed the content of the corresponding experience if remembering is not constructive. But the second does not entail the first: constructive remembering might in principle produce retrieved memories the content of which does not exceed the content of the corresponding experiences. “Preservationism”, in what follows, refers to the first idea. Note that preservationism is distinct from *transmissionism*, the view that memory traces transmit content from experience to retrieved representation (Hutto & Myin, 2017; Hutto & Peeters, 2018).

<sup>5</sup> We will see, however, that the simulationist argues that, while causalism may not entail preservationism, to reject preservationism is nevertheless to step onto a slippery slope that ends in the rejection of causalism.

<sup>6</sup> Because both refer to traces, both CTM and STM have been challenged on the ground that they presuppose transmissionism. It may be possible to formulate nontransmissionist versions of both causalism and simulationism (Michaelian & Sant’Anna online ahead of print; Werning 2020).

remembering is the retention of knowledge or justification (see Bernecker, 2008). Because the epistemic theory itself has largely been abandoned (though see James, 2017 for one recent defence), essentially on the ground that the notions of knowledge and justification are normative in character and thus out of place in an analysis of remembering, these challenges have little force. There have been challenges based on hybrid *epistemic-causal* (Debus, 2010) and *autonoetic-causal* (Dokic, 2014; Klein, 2015) theories to the claim that appropriate causation is (together with previous experience and current representation) sufficient for remembering. Because hybrid theories continue to accept the necessity of appropriate causation, they remain causal theories, and these challenges do not yet get at the heart of CTM. But there have also been challenges based on *postcausal* theories, which start from the same conceptual framework as causalism but claim that appropriate causation is not necessary for remembering. These challenges do get directly at the heart of CTM.

In addition to STM, the family of postcausal theories includes the functionalist theory of memory (Fernández, 2018, 2019). Because the functionalist theory has only recently been formulated, its implications for the relationship between remembering the past and imagining the future have just begun to be explored (Fernández, 2020), and its implications for the relationship between successful and unsuccessful remembering have yet to be investigated. The functionalist theory will therefore be set aside here; the remainder of the paper will focus on the causalist-simulationist dialectic rather than on the broader causalist-postcausalist dialectic. The basic strategy of the simulationist's argument has two steps: in the first step, he argues against the sufficiency of AC (together with PE and CR) for remembering; in the second step, he argues against the necessity of AC for remembering (Michaelian, 2016c).

The *anti-sufficiency* argument starts from the observation that empirical research on the constructive character of remembering<sup>7</sup> demonstrates that retrieving a memory of an event is not a matter of literally retrieving a stored representation deriving from one's experience of that event; it is, instead, a matter of generating a new representation on the basis of stored representations originating in multiple sources, including in one's experiences of multiple events. Though the constructive character of remembering does not imply that retrieved memories are necessarily inaccurate, it is best illustrated by the errors to which it sometimes gives rise. Consider, for example, memory conjunction errors (Reinitz & Lammers, 1992),<sup>8</sup> which occur when a subject studies stimuli composed of multiple features (e.g., faces composed of eyes, noses, mouths, hair, and so on) and later falsely recognizes a new stimulus composed of features drawn from different studied stimuli (e.g., a new face composed of the eyes and nose of one studied face and the mouth and hair of another studied face). There is evidence suggesting that conjunction errors occur for autobiographical events as well as laboratory material and that subjects judge that they

<sup>7</sup> Neither this research nor the research on memory as a form of mental time travel to which the anti-necessity argument appeals will be reviewed here; for reviews, see Addis (2018, 2020), Michaelian (2011, 2016c).

<sup>8</sup> This example is inspired by Danilo Fraga Dantas' talk at the Santa Maria-Grenoble Memory Workshop in 2019.

“remember” (as opposed to merely “knowing”) the new events (Odegard & Lampinen, 2004). This, in turn, suggests that retrieval often has a recombinatorial character, which in turn suggests that retrieved memories may sometimes be “appropriately” connected to multiple events (cf. Robins, 2017a).

Suppose, for example, that a subject S experiences events e1, composed of features f1, f2, and f3, and e2, composed of features f1, f2, and f4. (For instance, e1 might be the event of greeting a colleague, who was wearing a certain pair of glasses (f1) and a certain pair of shoes (f2), on Monday morning (f3), and e2 might be the event of greeting the same colleague, who was wearing the same pair of glasses (f1) and the same pair of shoes (f2), on Tuesday morning (f4).) Suppose that S later retrieves a memory m that represents an event composed of f1 and f2 (and is silent about f3 and f4) and that the component of m’s content that represents f1 derives from a memory trace originating in S’s experience of e1, while the component of m’s content that represents f2 derives from S’s experience of e2. AC is satisfied by m both with respect to e1 and with respect to e2, as are PE and CR.<sup>9</sup> CTM thus implies that, when S retrieves m, he remembers both e1 and e2.

Now, it may be possible for a given memory to be a memory of multiple events. But given that retrieval is frequently recombinatorial, CTM would seem to imply that retrieved memories are very often of multiple events. The implausibility of this implication means that research on the constructive character of remembering—and in particular the finding that retrieval is a matter of generating a representation on the basis of stored representations originating in experiences of multiple events—suggests that appropriate causation is not (together with previous experience and current representation) sufficient for remembering.

The core idea of the *anti-necessity* argument is that empirical research on the constructive character of remembering and on memory as a form of mental time travel suggests that appropriate causation is not necessary for remembering. The simulationist argues, first, that the findings of research on the constructive character of remembering indicate that retrieved apparent memories routinely include new content (content not included in the corresponding earlier experiences). Such content may be generated by the subject at the time of remembering or may derive from other experiences or from sources such as testimony.<sup>10</sup> Given that this is a routine occurrence, we cannot maintain that retrieved apparent memories that include new content are not genuine memories, for doing so would commit us to maintaining that most apparent memories are not genuine. We must thus (rejecting preservationism) accept that remembering can occur in cases in which at least a minority of the content of the retrieved representation is new. At this point, however, we have stepped onto a slippery slope. If we accept that remembering can occur in cases in which only a minority of the content of the retrieved representation is new, there

<sup>9</sup> That PE is satisfied is built into the description of the case. Regarding CR, it is plausible that, if S represents one of e1 and e2, he represents the other. While one might in principle maintain that S represents neither e1 nor e2, it is more natural, given the description of the case, to take him to represent both.

<sup>10</sup> See the discussion of the “lost in the mall” paradigm and the misinformation effect in Sects. 3 and 4 below.

is no non-arbitrary reason to deny that it can occur in cases in which a majority of the content of the representation is new. And if we accept that remembering can occur in cases in which a majority of the content of the representation is new, there is no non-arbitrary reason to deny that it can occur in cases in which the entirety of the content is new. But if we grant that remembering can occur in cases in which the entirety of the content of the retrieved representation is new, we have in effect rejected AC, for then we can no longer require that the causal connection between the current representation and the earlier experience (if there is one) be sustained by a memory trace.

The simulationist argues, second, that a multitude of empirical findings from research on memory as mental time travel demonstrates that episodic memory and forms of imagination such as episodic future thought have similar phenomenologies, involve similar cognitive processes, and rely on similar brain regions. These findings, he continues, suggest that the attempt to distinguish between remembering the past and imagining it—in the imagination-as-process sense—is ultimately misguided: episodic memory is a form of episodic imagination, underwritten by the same episodic construction system as episodic future thought and other forms of episodic imagination. Episodic future thinking is a matter of imaginatively projecting oneself into the personal future. It could not and therefore does not involve an appropriate causal connection between the imagined event and the subject's representation of it. Episodic remembering, as the simulationist conceives of it, is a matter of imaginatively projecting oneself into the personal past. It may sometimes but—given its relationship to episodic future thinking—need not always involve an appropriate causal connection between the remembered event and the subject's representation of it. The simulationist thus takes research on the constructive character of remembering and on memory as a form of mental time travel to suggest that appropriate causation is not necessary for remembering.<sup>11</sup>

As far as the distinction between remembering the past and imagining it—in the imagination-as-failure sense—is concerned, the simulationist argues, as we will see in Sect. 3, that what makes the difference is the *reliability* of the process. The simulation theory can thus be formulated as follows (Michaelian, 2016c).<sup>12</sup>

---

<sup>11</sup> The causalist might object here that similar phenomenologies might be produced by different processes and that the phenomenological similarity between remembering the past and imagining the future should thus not be taken to imply that, since appropriate causation is not necessary for episodic future thinking, it is not necessary for episodic remembering. The simulationist's argument, however, appeals to similarities between remembering the past and imagining the future not only at the phenomenological but also at the cognitive and neural levels. While it is open to the causalist to argue that even such extensive similarity between remembering the past and imagining the future does not demonstrate that appropriate causation is not necessary for episodic remembering, responding to such an argument would take us too far afield here. See Perrin and Michaelian (2017) for discussion of the options that are open to the causalist at this stage of the dialectic.

<sup>12</sup> While Shanton and Goldman (2010) and De Brigard (2014a, b) share many of the inspirations and concerns of Michaelian 2016c, they are silent with respect to AC. The focus in what follows will be on Michaelian's version of simulationism, but De Brigard's will also be discussed.



- (STM) S now remembers an event *e* iff  
 (CR) S now represents *e*;  
 (R) S's current representation of *e* is produced by a properly functioning and hence reliable episodic construction system that aims to produce a representation of an event belonging to S's personal past.

STM is compatible with the claim that remembering *sometimes* involves an appropriate causal connection. But it parts ways with CTM by rejecting the claim that remembering *always* involves an appropriate causal connection. STM and CTM will thus issue conflicting verdicts about cases in which reliability and appropriate causation come apart (as they do, for example, in the “lost in the mall” cases discussed below).

### 3 Confabulating as unreliable imagining

Like CTM, STM has been challenged in a variety of ways. There are, as noted above, challenges based on apparent differences between episodic memory and episodic future thought. These challenges, which have led to the *continuism-discontinuism* debate, will not be discussed here. There are challenges suggesting that, because STM borrows the concept of reliability that figures in R from reliabilist epistemology, it amounts to an *epistemic theory* of memory (Bernecker, 2017). These challenges are indirectly relevant to the concerns of the paper and will be discussed briefly in Sect. 4. And there are challenges suggesting that, because STM treats remembering as a form of imagining, it is unable to distinguish between successful remembering and forms of unsuccessful remembering or mere imagining such as confabulating (Robins, 2016a). These challenges and the *memory error debate* that they have triggered are the focus of the present section.<sup>13</sup>

#### 3.1 The memory error debate

The importance of the memory error debate is best brought out by considering the overall shape of the causalist-simulationist dialectic. As we saw in Sect. 2, causalists argue, on the basis of intuition, for the necessity of appropriate causation, while simulationists argue, on the basis of empirical evidence, against the necessity of appropriate causation. Positive empirical evidence for the claim that, in each and every case of remembering, there is a causal connection, sustained by a memory originating in the experience, between the subject's current representation of the event and his previous experience of it is difficult to come by. To the extent that they are concerned with formulating an empirically plausible version of CTM, causalists have thus been led to adopt a defensive posture.

<sup>13</sup> More recent challenges to which simulationists will eventually need to respond pertain to the concept of simulation (Andonovski 2019), the objects of memory (Aranyosi 2020), and the function of memory (Schwartz 2020).

Consider Werning's (2020) minimal trace theory. Werning holds a view of memory traces very unlike that held by Martin and Deutscher, but he continues to maintain that, in each and every case of remembering, there is an appropriate causal connection between the subject's current representation of the event and his previous experience of it. Because he recognizes the constructive character of remembering, and in particular because he grants that memories are in some cases based not on experiences of the events that they represent but rather on testimony about those events, however, he is able to maintain this only by claiming that what he refers to as "vicarious experiences" are genuine experiences.

Werning is led to make this claim because he takes *vicarious memories* (Pillemer, Steiner, Kuwabara, Thomsen, & Svob, 2015) to be genuine memories, where vicarious memories are memories of events in which the rememberer was not involved but was, rather, told about but that are indistinguishable, from the rememberer's point of view, from memories of events in which he was involved. The structure of the sort of case that Werning has in mind is the following: S1 experiences event  $e_1$ ; S1 tells subject S2 about  $e_1$ ; upon hearing about  $e_1$ , S2 imagines  $e_1$  in detail; S2's memory system stores a trace deriving from his imagining of  $e_1$ ; this trace is later retrieved, resulting in the formation of a representation of  $e_1$  that S2 takes to be a memory. Along the same lines, Michaelian (2016c) describes a case of childhood amnesia having the following structure: S1 and S2 experience  $e_1$ ; S2 loses all memory of  $e_1$ ; S1 tells S2 about  $e_1$ ; upon hearing about  $e_1$ , S2 imagines it in detail; S2's memory system stores a trace deriving from his imagining of  $e_1$ ; this trace is later retrieved, resulting in the formation of a representation of  $e_1$  that S2 takes to be a memory. What the vicarious memory case and the childhood memory case have in common is that, though there is a causal connection between S2's current representation of  $e_1$  and  $e_1$  in both cases, that connection is not sustained by a memory trace originating in S2's experience of  $e_1$ : "the causal route goes via another person ... and the information about the event to be remembered is conveyed by a verbal report" (Werning, 2020, p. 324). The friend case by means of which Martin and Deutscher originally motivated AC has the same feature.

Because Werning disagrees with Martin and Deutscher and agrees with Michaelian in affirming that cases having this feature may be cases of remembering, he is able to endorse causalism—to maintain that, in each and every case of remembering, there is a causal connection, sustained by a memory trace originating in the experience, between the subject's current representation of the event and his previous experience of it—only by claiming that, in cases of this kind, when the rememberer imagines an event on the basis of the testimony that he has received, he thereby experiences it: "trace minimalists", he says, "will allow vicarious experiences to play the role of experiences in their account of episodic memory". The term "vicarious experience" is, however, misleading. What Werning refers to as a vicarious experience certainly is *an experience*, as is any imagining. Just as certainly, however, it is not an experience *of the imagined event*. Once this is granted, it becomes clear that Werning's causalism is a causalism in name only. In the vicarious memory case, S2 does not experience the event when it occurs; it is his subsequent imagination, based on S1's experience of the event, that serves as the basis of the memory. In the childhood amnesia case, S2 does experience the event when it

occurs, but his experience of it does not serve as the basis of his eventual retrieved memory of it; instead, it is his subsequent imagination, based on S1's experience of the event, that serves as the basis of the memory. A simulationist might accept that S2 remembers in these cases.<sup>14</sup> But a causalist cannot coherently accept that S2 remembers in either case, for neither case involves an appropriate causal connection between S2's apparent memory of *e1* and his original experience of *e1*.

In light of the difficulties faced by causalists in formulating an empirically plausible version of CTM, the memory error debate takes on a particular importance: if causalist challenges suggesting that STM is unable to distinguish between successful remembering and forms of unsuccessful remembering such as confabulating were to succeed, they would undermine simulationism on the empirical terrain on which it is most at home, thereby reinforcing the empirical credentials of causalism. The issue at the heart of the debate is the nature of confabulation and its relationship both to remembering and to related errors. Confabulation can, roughly, be defined as an error in which a subject who is unable to remember instead “makes up” a past event, either through the “dislocation of events in time” or through the “fabrication of stories to fill in forgotten sequences” (Goodwin, 1989, p. 65). Korsakoff, for example, describes a patient who, “[w]hen asked to tell how he has been spending his time ... would very frequently relate a story altogether different from that which actually occurred, for example, he would tell that yesterday he took a bike ride into town, whereas in fact he has been in bed for two months, or he would tell of conversations which never occurred” (1889/1955, p. 399; quoted at Robins, 2019, p. 2136). The rough definition and the anecdote serve to emphasize that confabulation has historically been understood as an error specifically of *memory* that occurs specifically in *clinical* contexts. It may be similar in certain respects to clinical but nonmnemonic errors such as delusion and to mnemonic but everyday errors such as misremembering, but it differs fundamentally from these in that it presupposes a breakdown or malfunction of the memory system, such as occurs in disorders such as amnesia. Mercer et al. are explicit on this point, claiming that “a necessary prerequisite for confabulation is impaired memory” (1977; quoted at Berrios, 1998, p. 226), as is Moscovitch, who claims that “[when] neuronal structures involved in the reconstructive [memory] process are damaged, memory distortion becomes prominent and results in confabulation” (1995; quoted at Berrios, 1998, p. 227).

Though confabulators' memory reports are sometimes intrinsically implausible, they are often plausible when considered in isolation from the context in which the subject makes them. Consider the patient described by Korsakoff, or Dalla Barba's patient MB, who, “while he was hospitalized, said on one occasion that he was looking forward to the end of the testing session because he had to go to the general

---

<sup>14</sup> Despite Werning's suggestion to the contrary, simulationism does not in fact count vicarious memories as genuine memories: STM says that a subject remembers an event only if the event that he apparently remembers belongs to his personal past, which implies, assuming that a subject's personal past includes only events in which he was involved, that one cannot remember an event in which one was not involved. The restriction of memory to cases in which the represented event belongs to the subject's personal past may, however, ultimately turn out to be inconsistent with the spirit of simulationism, and the simulationist ought therefore to take vicarious memory cases seriously.

store to buy some new clothes, since he hadn't been able to the day before, because he had gotten lost in the center of Paris, where he had fortunately met a nurse who kindly took him back to the hospital" (2009, p. 227).<sup>15</sup> Though not especially likely, the events described by the patient are plausible enough. Both when confabulators' memory reports are plausible and when they are implausible, they are the result of a malfunctioning memory system and thus tend to bear some relationship, however loose, to the subjects' actual pasts. Consider Dalla Barba's patient SD, who when asked what he did yesterday, replied: "Yesterday I won a running race and I have been awarded with a piece of meat which was put on my right knee" (2002, p. 227). Though extremely implausible, the event described by the patient is in fact composed largely of elements drawn from his personal past: SD, Dalla Barba tells us, "was actually involved in running races", and "[i]t was actually during a running race in the mountains that he fell, sustaining a severe head trauma and an open wound to his right knee" (227).

While the definitions of confabulation offered by empirical memory researchers are a step in the right direction, they make *falsity* or inaccuracy into an essential feature of confabulation (see, e.g., Dalla Barba, 2002, 2009, 2016). As Hirstein (2005) and Robins (2016b) have pointed out, veridical confabulation is possible even if improbable. These definitions are therefore inadequate from a philosophical point of view and have found no advocates in the current debate. An earlier philosophical account, the *epistemic* account (Hirstein, 2005, 2011), according to which confabulating is distinguished from remembering by the unjustifiedness of the apparent memories to which the process gives rise, recognizes the possibility of veridical confabulation. The epistemic account, however, presupposes the epistemic theory of memory, which, as noted above, has largely been abandoned. This account has therefore likewise found no advocates in the current debate. The current debate has thus unfolded between partisans of rival causalist (Bernecker, 2017; Robins, 2016a, 2019, 2020a, b) and simulationist (Michaelian, 2016b, 2020) accounts of confabulation.

### 3.2 The causalist classification of memory errors

The memory error debate was triggered by Robins' claim that, because it treats remembering as a form of imagining, simulationism is unable to distinguish between successful remembering and forms of unsuccessful remembering such as confabulating. Picking up on De Brigard's suggestion that, both in cases of remembering and in cases of imagining, the memory system "is doing what it is supposed to do" (2014a, p. 172), Robins argues that, while simulationism is able to take the potential inaccuracy of retrieved apparent memories into account, it is unable to take potentially problematic features of the retrieval process into account and is therefore unable to distinguish between remembering and confabulating. Simulationism, as

<sup>15</sup> Note that MB's claim concerns both his personal *past* and his personal *future*. Dalla Barba (2002, 2009, 2016) has consistently emphasized the possibility of future-oriented confabulation, but this possibility has yet to be discussed in the current debate; see Sect. 5 below.

she sees the matter, “collapse[s] the processing distinction between memory errors and successful remembering” (2016a, p. 441).<sup>16</sup>

Causalism does better than simulationism with respect to unsuccessful remembering, in Robins’ view, simply because it is able to appeal to appropriate causation. Robins (2016a) proposes a classification of successful remembering and forms of unsuccessful remembering based on two conditions. The first is an *accuracy* condition, which requires that the subject form an accurate representation of the past event. The second is a *retention or appropriate causation* condition equivalent to AC, which requires that the subject’s representation be based on retained information originating in his experience of the event. In *successful remembering*, Robins argues, both conditions are satisfied. In *confabulating*, neither condition is satisfied. Confabulations are thus characterized both in terms of their falsity and in terms of their failure to be causally connected, via a memory trace, to corresponding earlier experiences. In addition to confabulating, Robins’ classification recognizes two errors. In *relearning*—as seen in the friend and childhood amnesia cases described above—the accuracy condition is satisfied but the appropriate causation condition is not. In *misremembering*—an everyday memory error typified by the DRM effect, in which the subject is presented with a list of thematically-related words (e.g., *hospital, sick, nurse ...*) and later recalls having seen a thematically-consistent but non-presented lure word (e.g., *doctor*) (Gallo, 2010)—the appropriate causation condition is satisfied but the accuracy condition is not.

### 3.3 The simulationist classification of memory errors

Robins’ claim that, because simulationism treats remembering as a form of imagining, it is unable to distinguish between remembering and confabulating has led to an ongoing debate over the abilities of causalism and simulationism to generate appropriate classifications of memory errors (see Bernecker forthcoming, Michaelian forthcoming).

#### 3.3.1 The initial simulationist classification

In response to Robins, Michaelian (2016b) proposes two classifications, one of which includes an internality condition designed to acknowledge relearning by requiring that the subject himself contribute content to the content of the retrieved representation. It is unclear, however, whether relearning, in which the causal route

<sup>16</sup> Robins’ argument may succeed against De Brigard’s version of simulationism, which, unlike Michaelian’s, does not include a reliability condition. (See Robins 2019 for discussion.) Whereas Michaelian emphasizes the relationship between episodic memory and episodic future thought, De Brigard emphasizes the relationship between episodic memory and episodic counterfactual thought. It may be natural, given an emphasis on the relationship between memory and counterfactual thought, not to include a reliability condition in one’s theory of memory, but this is not inevitable. If a condition along the lines of R were to be incorporated into De Brigard’s version of simulationism, a De Brigard-style simulationist should be able to avail himself of an approach to confabulation along the lines of that defended by Michaelian.

between the event and the apparent memory “goes via another person” (Werning, 2020, p. 324), should ultimately be counted as a memory error (Bernecker, 2017). In relearning cases, the subject may take his apparent memory, which is in fact based on testimony, to be based on his own experience of the remembered event, but he need not do so. If he does do so, moreover, the error that he commits is a source monitoring error (Johnson, Hashtroudi & Lindsay 1993; see Michaelian, 2012), an error of a sort other than those with which the memory error debate is concerned. Michaelian’s treatment of relearning will thus be set aside here.

It is clear that misremembering, in contrast to relearning, should be counted as a memory error, and an important advantage of Robins’ classification is that it accommodates this error. An important disadvantage of her classification is that it fails to accommodate veridical in addition to falsidical confabulation. The second classification proposed by Michaelian is designed to accommodate veridical confabulation. Like that proposed by Robins, this classification, depicted in Table 1, is based on two conditions. The first is an *accuracy* condition equivalent to Robins’. The second is a *reliability* condition equivalent to R. The inclusion of R in STM is, indeed, inspired by the thought that the central difference between remembering and confabulating is the manifest unreliability of a retrieval process characterized by the “dislocation of events in time” and the “fabrication of stories to fill in forgotten sequences” (Goodwin, 1989, p. 65). In *successful remembering*, both conditions are satisfied. In *falsidical confabulation*, neither condition is satisfied. In *veridical confabulation*—in which the subject’s representation is accurate despite being confabulatory—the accuracy condition is satisfied but the reliability condition is not. In *misremembering*, the reliability condition is satisfied but the accuracy condition is not. Just as falsidical and veridical confabulation are united by the unreliability of the process that produces them, successful remembering and misremembering are united by the reliability of the process that produces them.

The picture proposed by the simulationist can thus be summed up by saying that what makes the difference between *genuine* and *merely apparent* remembering is reliability, whereas what makes the difference between *successful* and *unsuccessful* remembering is reliability plus accuracy. The causalist could in principle propose a parallel picture. (Indeed, Robins (2020b) has revised her classification so as to take veridical confabulation into account.) Such a picture would, however, not amount to a particularly faithful portrait of the cases of confabulation reviewed above or of the countless similar cases described in the literature. As noted above, confabulation is standardly understood as a clinical phenomenon, in that it presupposes a malfunction of the memory or episodic construction system. Given that appropriate causation might be absent even when the system functions properly—it is, as we have seen, absent in relearning—the concept of reliability would seem to be better suited to capturing the notion of confabulation than is the concept of appropriate causation.

Robins’ preference for understanding confabulation in causal terms may be due to the fact that she works with a nonstandard notion of confabulation, a notion on which at least some confabulations are everyday rather than clinical phenomena. “Suggestibility studies”, she writes, citing Loftus and Pickrell (1995) (see also Loftus, 1996), “are examples of confabulation. These studies show that, as a result of mildly suggestive questioning, participants can come to ‘remember’ events they

never experienced, such as being lost in a shopping mall as a small child or having been hospitalized overnight” (2016a, p. 434; cf. Robins, 2020b). If confabulation includes everyday errors such as those that occur in “lost in the mall” (LITM) cases, it does not presuppose a malfunction of the system and thus cannot be defined in terms of unreliability; it might, however, be feasible to understand it in causal terms.

It is, however, a mistake to stretch the notion of confabulation to include everyday errors. French, Garry, and Loftus, for example, treat everyday false memories as confabulations (acknowledging that in doing so they depart from the standard understanding of confabulation) and argue that “such confabulations [in non-clinical subjects] are a byproduct of normally functioning memory processes and mechanisms” (2009, p. 34). But there is a clear and important difference between a subject in a psychology experiment who, having received suggestions that he was once lost in the mall, comes to believe that he was once lost in the mall and a subject who, when asked how he has been spending his time, volunteers that “yesterday he took a bike ride into town” (Korsakoff 1889/1955, p. 399), despite having been in bed for two months or who says that “[y]esterday [he] won a running race and [was] awarded with a piece of meat which was put on [his] right knee” (Dalla Barba, 2009, p. 227). Indeed, there are two clear and important differences. First, the underlying mechanisms are different: the participants in Loftus’s experiments have properly functioning episodic construction systems, whereas Korsakoff’s and Dalla Barba’s patients do not. Second, the participants in Loftus’s experiments produce inaccurate apparent memories relatively infrequently, whereas Korsakoff’s and Dalla Barba’s patients produce them much more frequently. These two differences make clear that, while LITM cases are cases of error, the error in question is of a different kind than the error of confabulation. It is thus misleading to apply the term “confabulation” both to the clinical errors and to the everyday errors.

Of course, even if LITM cases and other cases of “everyday confabulation” are not in fact instances of confabulation, the simulationist must nevertheless decide how to classify them. Indeed, LITM cases illustrate the way in which causalist and simulationist approaches issue conflicting verdicts in cases in which reliability and appropriate causation come apart: in an ordinary LITM case in which the reliability condition is satisfied, simulationism will imply that the subject is misremembering, whereas causalism implies that the subject is confabulating. The simulationist treatment of other kinds of LITM case, as well as cases involving the misinformation effect, the investigation of which was also pioneered by Loftus (2005), is discussed in Sect. 4.

The point that the participants in Loftus’s experiments produce inaccurate apparent memories relatively infrequently, whereas Korsakoff’s and Dalla Barba’s patients produce them much more frequently, should not be taken to mean that simulationism distinguishes between confabulating and misremembering in terms of their respective frequencies, as Robins (2019) suggests. Though it is highly plausible that clinical subjects remember unsuccessfully more frequently than do non-clinical subjects, the frequency of confabulation and misremembering is an empirical matter, and it would be a problem for simulationism were it to imply that clinical subjects necessarily remember unsuccessfully more frequently than do non-clinical subjects. Simulationism does not, however, imply this. What STM’s reliability condition, *R*, says

**Table 1** The simulationist classification (Michaelian, 2016a, b, c)

$R$		$\sim R$	
$A$	$\sim A$	$A$	$\sim A$
remembering	mis-remembering	veridical confabulation	falsidical confabulation

is that the representation in question was produced by a properly functioning and hence reliable episodic construction system—that is, by a system that functioned properly and hence reliably *when it produced the representation in question*. That  $R$  is (not) satisfied on a particular occasion on which a subject apparently remembers thus does not imply that it is (not) satisfied on other occasions on which that subject apparently remembers, simply because a system that functions properly on one occasion may malfunction on another and a system that malfunctions on one occasion may function properly on another. The simulationist approach is thus compatible with the possibility that a subject who usually (mis)remembers occasionally (falsidically or veridically) confabulates and the possibility that a subject who usually (falsidically or veridically) confabulates occasionally (mis)remembers. It is likely, in fact, that, in many clinical conditions that are characterized by confabulation, subjects have episodic construction systems that sometimes malfunction but sometimes function properly.<sup>17</sup>

### 3.3.2 The revised simulationist classification

Overall, then, simulationism appears to do better than causalism with respect to confabulation. The simulationist picture presented above, however, does not yet provide a fully faithful portrait of the phenomenon, for it fails to take the role of *metamemory* into account. We saw above that a certain form of metamemory error, namely, source monitoring error, may be at work in some cases of relearning. Another form of metamemory error appears to be at work in paradigm cases of confabulating, for such cases have two characteristic dimensions. Paradigm cases of confabulating involve, first, the production of (mostly) inaccurate representations due to mechanisms such as the “dislocation of events in time” and the “fabrication of stories to fill in forgotten sequences” (Goodwin, 1989, p. 65). They involve, second, a failure, on the part of the subject, to recognize, even in cases in which the resulting representations are highly implausible or incongruous with reality, that something has gone wrong with the retrieval process and hence to reject the representations in question. As Schnider puts it, confabulators characteristically “[fail] to suppress—or

<sup>17</sup> Causalists will no doubt be unconvinced by this brief argument against their treatment of unsuccessful remembering and in favour of the simulationist treatment. See Michaelian 2020 for a more detailed argument.



rather filter—activated memory traces and mental associations which do not refer to current reality” (2018, p. 215; cf. Hirstein, 2005).

The simulationist classification considered above focuses exclusively on the first dimension. If the second dimension is taken into account, confabulating would seem, from a simulationist perspective, to involve both *object-level* unreliability, resulting, in most cases, in the production of an inaccurate representation, and *meta-level* unreliability, resulting, in most cases, in a failure to detect the object-level unreliability. Note that metacognitive monitoring processes cannot directly detect the accuracy of a retrieved apparent memory but can detect features of the retrieval process (e.g., fluency) that are correlated with its reliability and hence with the probable accuracy of the apparent memories that it produces; the accuracy of metacognitive judgements is thus to be understood as accuracy with respect to the reliability of the relevant retrieval process.

It is crucial to note that reliability and accuracy can come apart both at the object level and at the meta level, resulting in a form of *luck*. When they come apart at the object level, the result is either an unluckily inaccurate apparent memory (misremembering) or a luckily accurate apparent memory (veridical confabulation). When they come apart at the meta level, the result is either an unluckily inaccurate metacognitive judgement or a luckily accurate metacognitive judgement. This form of luck plays an important role in the revised simulationist classification of successful remembering and forms of unsuccessful remembering proposed by Michaelian (2020).<sup>18</sup>

The revised classification, depicted in Table 2, includes object-level reliability and accuracy conditions but supplements these with meta-level reliability and accuracy conditions. On the revised classification, *successful remembering* occurs when both conditions are satisfied at both levels. It thus involves luck at neither level: at the object level, a reliable process produces an accurate representation of a past event; at the meta level, a reliable process produces an accurate evaluation of the reliability of the object-level process. *Falsidical confabulating* occurs when neither condition is satisfied at either level. It thus likewise involves luck at neither level: at the object level, an unreliable process produces an inaccurate representation of a past event; at the meta level, an unreliable process produces an inaccurate evaluation of the reliability of the object-level process. *Misremembering* occurs when the reliability condition is satisfied at both levels and the accuracy condition is satisfied at the meta level but not the object level. It thus involves luck at the object level but not the meta level: at the object level, a reliable process happens to produce an inaccurate representation of a past event; at the meta level, a reliable process produces an accurate evaluation of the reliability of the object-level process. *Veridical confabulating* occurs when the reliability condition is satisfied at neither level and

---

<sup>18</sup> The notion of luck at work here pertains not to reliability (i.e., luck does not imply unreliability) but rather to the relationship between reliability and accuracy. While the discussion of Sect. 4 will show that this needs to be refined further, luck can provisionally be understood as being at work when the memory process does not produce the “expected” result, that is, when a reliable process produces an inaccurate representation or when an unreliable process produces an accurate representation.

the accuracy condition is satisfied at the object level but not the meta level. It thus likewise involves luck at the object level but not the meta level: at the object level, an unreliable process happens to produce an accurate representation; at the meta-level, an unreliable process produces an inaccurate evaluation of the reliability of the object-level process.

A key advantage of the revised classification is that it brings out the fact that there are many ways, in addition to misremembering and (veridical and falsidical) confabulation, in which attempted remembering can go wrong. One might think here of Aristotle's observation that there are many ways to fail but only one way to succeed, a thought to which we will return in Sect. 4. Some of these errors involve no luck; others involve either object-level luck or meta-level luck; others still involve both object-level and meta-level luck. These more exotic errors (discussed in detail in Michaelian, 2020) might occur only infrequently and might be difficult to detect when they do occur. In what we might refer to as "innocently-endorsed veridical confabulation", for example, which occurs when, at the object level, an unreliable process happens to produce an accurate representation of a past event and, at the meta level, a reliable process happens to produce an inaccurate evaluation of the reliability of the object-level process, with the result that the subject endorses the retrieved representation, the subject ends up with an accurate apparent memory, as good luck and bad luck cancel each other out in a structure reminiscent of that of Gettier cases.<sup>19</sup> But that does not mean that they are unimportant. Gettier cases, while similarly infrequent and difficult to detect, have nevertheless taught us much about the nature of knowledge. These exotic errors may similarly have much to teach us about the nature of memory.

#### 4 Remembering as apt imagining

One thing that they may tell us is that STM is inadequate as it stands. The initial simulationist classification already acknowledged errors involving object-level luck. The revised simulationist classification additionally acknowledges errors involving meta-level luck. STM itself, however—the theory, as opposed to the classification—takes neither the meta level itself nor the possibility of luck—whether at the object level or at the meta level—into account. The theory thus stands in need of revision. Aristotle's observation, made in the course of his analysis of virtuous action, that there are many ways to fail but only one way to succeed, suggests that the way forward may lie in the development of a *virtue theory of memory*. The possibility of moving to a virtue theory of memory is likewise suggested by reflection on the relationship between theories of memory and theories of knowledge.

Bernecker (2017) claims that the simulationist account of confabulation, which distinguishes between (mis)remembering and (veridical and falsidical) confabulating in terms of the (un)reliability of the episodic construction system, is an epistemic

---

<sup>19</sup> The analogy between innocently-endorsed veridical confabulation and Gettier cases is imperfect; a memory error more closely analogous to Gettier cases is discussed in Sect. 4.2.

**Table 2** The revised simulationist classification (Michaelian, 2020). Light grey cells indicate object-level luck. Mid-grey cells indicate meta-level luck. Dark grey cells indicate both object-level and meta-level luck

		<i>object level</i>				
		<i>R</i>		<i>~R</i>		
		<i>A</i>	<i>~A</i>	<i>A</i>	<i>~A</i>	
<i>meta level</i>	<i>R</i>	<i>A</i>	successful remembering	mis-remembering		
		<i>~A</i>				
	<i>~R</i>	<i>A</i>				
		<i>~A</i>			veridical confabulating	falsidical confabulating

account, implying that STM is an epistemic theory—specifically, a reliabilist theory—of memory. Michaelian (2020) points out, in response, that this claim is mistaken. Compare the simulationist account of confabulating to that proposed by Hirstein (2005), which treats confabulating as “ill-grounded”. Because he makes use of epistemic concepts in his account of confabulating, Hirstein would seem to be committed to an epistemic theory of memory. Simulationism, in contrast, treats confabulating as unreliable. Because reliability is not an epistemic concept, the simulationist is not committed to an epistemic theory of memory. It may nevertheless be useful, for heuristic purposes, to take the analogy between the simulation theory of memory and the reliabilist theory of knowledge seriously. Indeed, there is a potential analogy between the development of the externalist family of theories of knowledge of which reliabilism is a member and the development of the family of (post)causal theories of memory of which simulationism is a member. Simplifying greatly, we might say that the limitations of the causal theory of knowledge (Goldman, 1967) led to the emergence of the process reliabilist theory of knowledge (Goldman, 1979, 2012) and that the limitations of process reliabilism led reliabilists to move to the virtue reliabilist theory of knowledge (Pritchard, 2012; Sosa, 2007, 2009).<sup>20</sup> Along the same lines, we might say that the limitations of the causal theory of memory led to the emergence of the simulation theory of memory (a process reliability theory) and that the limitations of simulationism suggest that simulationists move to a virtue reliability theory of memory.

Bearing this analogy in mind, the key insight deriving from the memory error debate would seem to be that there is a form of *mnemonic luck* that is absent in successful remembering and present in many kinds of unsuccessful remembering, a form of luck analogous to the form of *epistemic luck* the existence of which was

<sup>20</sup> This description of the emergence of virtue epistemology simplifies in part by leaving out non-reliabilist virtue approaches. It would be worth considering whether a space of potential reliabilist and non-reliabilist virtue approaches to memory, broadly analogous to the space of reliabilist and non-reliabilist virtue approaches to knowledge explored by Battaly (2008), might be mapped out.

brought to light by the literature on Gettier cases, which occur when a justified belief is only luckily true and therefore fails to amount to knowledge (Pritchard & Smith, 2004; Zagzebski, 1994). The existence of mnemonic luck may have been overlooked until now simply because the influence of the epistemic theory has waned as that of the causal theory has waxed: as they have been, for the most part, uninterested in the epistemic theory, philosophers of memory have not yet considered whether and how insights from recent analyses of knowledge—and in particular from recent virtue epistemologies, which are expressly designed to cope with epistemic luck—might be applied to the analysis of memory.<sup>21</sup>

Virtue epistemology treats knowledge as an *achievement*, in the sense that, in order for a true belief to qualify as knowledge, its truth must be due to the subject's cognitive capacities, ruling out the involvement of luck. Taking virtue epistemology—in particular, Sosa's virtue theory—as a model, the remainder of this section develops a virtue theory of memory designed to acknowledge the absence of luck in successful remembering and its presence in many kinds of unsuccessful remembering. The virtue theory of memory and the virtue-theoretic or aretaic classification of memory errors that it suggests, in turn, entail a reconceptualization of memory as an achievement.

#### 4.1 The virtue theory of memory

It will simplify matters to bracket the meta level, coming back to it once we have the workings of luck at the object level clearly in view.

##### 4.1.1 Object-level virtue

The revised simulationist classification of memory errors treats the joint satisfaction of the accuracy and reliability conditions as ruling out the presence of luck, but this is a mistake. Sosa's well-known archer metaphor vividly illustrates the point. Suppose that an archer hits his target. Suppose that he shot with skill. His hitting of the target might still be due to luck—his arrow might, for example, have been blown off course by an unexpected gust of wind (bad luck), only to be blown back on course by another gust of wind (good luck). Luck can be ruled out only if the target is hit *because* the arrow was shot with skill. In the domain of knowledge, similarly, epistemic luck can be ruled out only if a belief is true because it was formed by a reliable process. Consider Russell's well-known case of the stopped clock. Suppose that a subject truly believes that it is 9:00. Suppose that the subject formed this belief by looking at a clock that has always kept good time and believing what it indicated. His truly believing that it is 9:00 may still be due to luck—the clock might have stopped at 9:00 the day before (bad luck), but the subject might have happened to look at it at precisely 9:00 today (good luck). In the domain of memory, mnemonic

---

<sup>21</sup> James (2017) is an important exception.

luck can be ruled out only if an apparent memory is accurate because it was produced by a reliable retrieval process. Consider what we might describe as a “lucky LITM” case.<sup>22</sup> Suppose that a subject has an accurate apparent memory of being lost in the mall as a child. Suppose that the retrieval process that produced this apparent memory was reliable. The accuracy of the apparent memory may still be due to luck—the subject might, for example, have been the victim of experimenters seeking to implant in him a false memory of being lost in the mall as a child (bad luck), but he might happen in fact to have been lost in the mall as a child (good luck).<sup>23</sup>

Sosa appeals to the fact that luck is possible despite accuracy and reliability to motivate his conception of knowledge as an achievement. An achievement, for Sosa is an *apt* performance, where aptness is understood as *accuracy* due to *adroitness*. A shot is apt if the arrow hits the target (accuracy) because the archer shoots with skill (adroitness). A belief is apt, for Sosa, if it is true (accuracy) because the relevant belief-forming capacity functions reliably (adroitness). Knowledge can thus be understood as apt belief. We might similarly appeal to this fact to motivate a virtue reliability theory of memory, a theory on which memory is an achievement. An apparent memory is apt, we might say, if it is accurate with respect to the relevant event (accuracy) because the episodic construction system functions reliably (adroitness). Successful memory can thus be understood as apt apparent memory.<sup>24</sup>

Unpacking the slogan that successful memory is apt apparent memory, a version of the virtue theory of memory that takes only the object level into account can be formulated as follows.

- (VTM<sub>O</sub>) S now remembers an event *e* iff
- (CR) S now represents *e*;
- (R) S’s current representation of *e* is produced by a properly functioning and hence reliable episodic construction system that aims to produce a representation of an event belonging to S’s personal past;
- (V) S’s current representation of *e* is accurate because it is produced by a reliable episodic construction system.

<sup>22</sup> See Michaelian forthcoming for a more detailed discussion of lucky LITM cases.

<sup>23</sup> This discussion makes do with an intuitive conception of luck, as will the remainder of the paper. Ultimately, of course, the virtue theory of memory will need to be supplemented with an explicit account of the nature of luck, but canvassing the available conceptions (see Hales 2020) and assessing their suitability for the virtue theory is not feasible here.

<sup>24</sup> Though the focus here is on reliabilist virtue theories, it should be noted that Turri (2015, 2016) has suggested that epistemologists should abandon reliabilism in favour of abilism, granting that knowledge is an achievement but arguing that achievements in general and knowledge in particular do not presuppose reliability (in the standard absolute sense) but only ability (in the sense of doing better than chance). If Turri is right, we might replace the virtue reliabilist theory of knowledge with a virtue abilist theory of knowledge, continuing to view knowledge as apt belief, continuing to understand aptness in terms of accuracy and adroitness, but now defining adroitness in terms of ability rather than reliability. Along the same lines, we might formulate a virtue ability theory of memory on which memory is apt apparent memory, where aptness is treated as accuracy due to adroitness and adroitness is treated as ability rather than reliability. A virtue ability theory of memory would likely require an overhaul of the simulationist classification of memory errors far more extensive than that entailed by the move from STM to the virtue reliability theory and undertaken below.

One might suspect that, even if STM was not an epistemic theory, its use of Sosa's "AAA" (aptness as accuracy due to adroitness) framework makes  $VTM_O$  an epistemic theory. But it is important to note that the AAA framework applies to any performance with an aim and that the aim in question need not be intentional: "[a] heartbeat", Sosa points out, "succeeds if it helps pump blood, even absent any intentional aim" (2007, p. 23). Thus, while the move to a virtue theory of memory is inspired by virtue epistemology, the adoption of the AAA framework does not turn simulationism into an epistemic theory of memory any more than its use of the concept of reliability turned it into an epistemic theory of memory. The virtue theory of memory would be an epistemic theory if it were to analyze memory in terms of knowledge, well-groundedness, or another epistemic concept, but it does not do so: just as reliability is not an epistemic concept (despite the fact that process and virtue reliabilists employ it in their analyses of the concept of knowledge), aptness is not an epistemic concept (despite the fact that virtue reliabilists employ it in their analysis of knowledge).

#### 4.1.2 Problems and clarifications

Conceiving of remembering as a performance with an aim nevertheless raises two problems. First, while some cases of remembering are intentional—voluntary or deliberate—others are not: they are involuntary or spontaneous. In cases of voluntary memory, it is plausible that the subject himself aims at forming an accurate representation of an event from his personal past. In cases of involuntary memory, however, the only thing that might aim at forming an accurate representation of an event from the subject's personal past is his episodic construction system, and it is unclear what it would be for an episodic construction system to *aim* at something. Second, assuming that we can make sense of the idea of an episodic construction system's aiming at something, there is room for disagreement over whether the episodic construction aims at forming an *accurate* representation of an event from the subject's personal past. The idea that the episodic construction aims at forming an accurate representation of an event from the subject's personal past can be challenged from two directions. On the one hand, De Brigard (2004) holds that memory aims not at forming representations of events that in fact occurred but rather at forming representations of events that might have occurred—memory, for him, is a form of hypothetical thought. On the other hand, Bernecker (2010; see also McCarroll, 2018) holds that memory aims not only at truth (where truth is accuracy with respect to the event itself) but also at authenticity (where authenticity is accuracy with respect to the subject's experience of the event).

Though these problems may become more obvious with the move to a virtue theory of memory, which explicitly conceives of remembering as a performance with an aim, they were already problems for STM, which required that the episodic construction system aim to produce a representation of an event belonging to the subject's personal past and thus implicitly conceived of remembering as a performance with an aim. The first problem remains open, but there is, given that it makes sense to understand other biological systems—such as hearts—as undertaking performances with aims, reason to suppose that it can be solved. The second problem likewise remains open, but there is reason to suppose that neither De

Brigard's nor Bernecker's view of the aim of remembering is correct. De Brigard's very undemanding view, Robins (2016a) argues, appears to imply that successful remembering is too frequent. And Bernecker's very demanding view, Michaelian and Sant'Anna (manuscript) argue, appears to imply that successful remembering is too rare. The view that the episodic construction system aims at forming an accurate representation of an event from the subject's personal past may thus be the golden mean between two untenable extremes.

Before turning to the meta level, four points about  $VTM_O$  (all of which also apply to  $VTM$ , stated below) should be noted. First,  $VTM_O$  explicitly refers to the accuracy of the subject's current representation. This is in contrast to the versions of  $CTM$  and  $STM$  given above, which do not refer to accuracy. Technically, then,  $CTM$  and  $STM$  are theories of genuine remembering, where genuine remembering includes both successful remembering and misremembering, whereas  $VTM_O$  is a theory of successful remembering. But this is simply a matter of convenience of expression: as we will see below,  $VTM_O$ , like  $CTM$  and  $STM$ , generates a classification of a full range of forms of unsuccessful remembering, including misremembering. One might worry that, despite the fact that  $CTM$  and  $STM$  are formulated as theories of genuine remembering (whether successful remembering or misremembering), the arguments for and against those theories reviewed in Sects. 2 and 3 presuppose that they are theories of successful remembering. The category of misremembering only came clearly into view with Robins, 2016a, and the arguments reviewed in Sect. 2 thus take accuracy for granted. Making this explicit does not seem to affect the success or failure of those arguments. And the arguments discussed in Sect. 3, which follow Robins, 2016a, explicitly invoke accuracy as the factor that makes the difference, within the category of genuine remembering, between successful remembering and misremembering.

Second,  $VTM_O$  is a virtue theory but remains a simulation theory, both in the sense that it rejects  $AC$  and in the sense that it treats remembering as a kind of imagining. The difference between  $VTM_O$  and  $STM$  is simply that, whereas  $STM$  says that to remember is to reliably imagine the past,  $VTM_O$ , because it includes the virtue condition ( $V$ ), says that to remember is to aptly imagine the past.

Third, because  $VTM_O$  includes  $V$ , its implications differ from those of  $STM$  with respect to certain cases. In particular, whereas  $STM$  implies that, in lucky LITM cases (and cases having a similar structure), the subject "simply remembers the episode" (Michaelian, 2016c, p. 119),  $VTM_O$  implies that the subject does not successfully remember in these cases. One might object here that it is not clear that lucky LITM cases ought to be treated as instances of unsuccessful remembering. Indeed, given that, in lucky LITM cases, there is an important sense in which nothing goes wrong with the memory process, one might maintain, more strongly, that they ought not to be treated as instances of unsuccessful remembering, in which case this point will count in favour of  $STM$  rather than  $VTM_O$ . Simulationists who are persuaded by this objection are free fall back to  $STM$ , but there is a strong case to be made for treating lucky LITM cases as instances of unsuccessful remembering. Consider again the archer metaphor. There is a sense in which nothing goes wrong with the shot: the arrow, after all, hits the target. There is, however, also a sense in which something goes wrong with it: the arrow's hitting the target is not attributable to the

archer's skill. It would clearly be a mistake to maintain that there is no difference worth nothing between this case and a luck-free case in which the arrow's hitting the target is attributable to the archer's skill. Similarly, there is a sense in which nothing goes wrong in a lucky LITM case: the retrieved memory is accurate. But there is also a sense in which something does go wrong: the memory's accuracy is not due to the proper functioning of the subject's episodic construction system. Just as it would be a mistake to overlook the difference between lucky and luck-free shots in archery, it would be a mistake to overlook lucky and luck-free accuracy in remembering. Granting this point, one might object that lucky LITM cases ought not to be treated as instances of unsuccessful remembering but rather as instances of successful remembering that are somehow defective. The objection might or might not be right, but the issue at this stage would seem to be merely verbal.

Fourth, V is not sufficient to rule out all forms of luck that we might in principle take to be incompatible with successful remembering. In the domain of knowledge, Pritchard (2012) has emphasized the importance of cases in which the subject's belief is true, produced by a reliable process, and true because it is produced by a reliable process but in which the belief is only luckily true. Consider the well-known "fake barn country" case. The subject looks out his car window, sees what appears to be a barn, and forms the belief that there is a barn. There really is a barn: his belief is true. And there is nothing unusual about his belief-forming process: the process was reliable, and his belief is true because the process was reliable. Nevertheless, the belief is only luckily true: the subject is (unbeknownst to him) driving through fake barn country, and the barn that he happened to look at is the only real barn in the vicinity. If he had looked at any other apparent barn, he would have ended up with a false belief. The possibility of *environmental epistemic luck* of this sort, Pritchard points out, means that aptness does not entirely rule out epistemic luck. If, as he holds, knowledge is incompatible with environmental epistemic luck (so that the subject in the fake barn case does not know that there is a barn), a virtue theory of knowledge like that defended by Sosa will be inadequate.

Cases having a structure like that of the fake barn case are possible in the domain of memory. Consider the following "environmental LITM" case. Attempting to remember an event that occurred in his early childhood, the subject retrieves an apparent memory of being lost in the mall. He really was lost in the mall: his apparent memory is accurate. And there is nothing unusual about his retrieval process: the process was reliable, and his apparent memory is accurate because the process was reliable. Nevertheless, the apparent memory is only luckily accurate: the subject has (unbeknownst to him) been participating in experiments designed to implant false childhood memories, and the event of being lost in the mall is the only event from his early childhood of which he is capable of retrieving an accurate apparent memory. If he had retrieved a memory of any other early-childhood event, he would have ended up with an inaccurate apparent memory.<sup>25</sup> There are thus cases in which

---

<sup>25</sup> Note that environmental LITM cases are not cases in which a subject who usually confabulates manages to remember on a particular occasion; they are cases in which a subject who usually remembers happens to find himself in circumstances in which he is likely to misremember but nevertheless ends up retrieving an accurate apparent memory.



the subject's apparent memory is accurate, produced by a reliable episodic construction system, and accurate because it is produced by a reliable episodic construction system but in which the memory is only luckily accurate. The possibility of *environmental mnemonic luck* of this sort means that aptness does not entirely rule out mnemonic luck. The question is whether memory is compatible with environmental mnemonic luck. If it is not,  $VTM_O$  will be inadequate.

Pritchard, because he holds that knowledge is not compatible with environmental luck, proposes an anti-luck virtue theory of knowledge, a virtue theory of knowledge supplemented by an explicit anti-luck condition. One who holds that memory is incompatible with environmental luck might similarly propose an anti-luck virtue theory of memory. Now, it is not entirely clear that knowledge is incompatible with environmental luck. Pritchard holds that it is not, but others, such as Sosa, hold that it is—that knowledge is “fragile”. It is even less clear whether memory is incompatible with epistemic luck: many achievements are fragile, and successful remembering may be among them. It would be premature to attempt to resolve this question here, but note that, because the virtue theory is not an epistemic theory, our judgement about whether memory is fragile may come apart from our judgement about whether knowledge is fragile. We might, in particular, agree with Pritchard that knowledge is not fragile, in which case the subject in an environmental LITM case does not *know* that he was lost in the mall, while maintaining that memory is fragile, in which case he does *remember* that he was lost in the mall.

#### 4.1.3 Meta-level virtue

Turning to the meta level, a version of the virtue theory of memory that takes both the object level and the meta level into account can be formulated as follows.

- (VTM) S now remembers an event  $e$  iff
- (CR) S now represents  $e$ ;
- (R) S's current representation of  $e$  is produced by a properly functioning and hence reliable episodic construction system that aims to produce a representation of an event belonging to S's personal past;
- (V) S's current representation of  $e$  is accurate because it is produced by a reliable episodic construction system;
- (CR-m) S accurately judges that his episodic construction system functioned reliably when it produced his representation of  $e$ ;
- (R-m) S's judgement that his episodic construction system functioned reliably when it produced his representation of  $e$  is produced by a reliable meta-memory monitoring process;
- (V-m) S's judgement that his episodic construction system functioned reliably when it produced his representation of  $e$  is accurate because it is produced by a reliable metamemory monitoring process.

The meta-level current representation, reliability, and virtue conditions (CR-m, R-m, and V-m) are simply the meta-level counterparts of CR, R, and V.

The rationale for the inclusion of conditions CR-m and R-m in the theory was given in Sect. 3.2.2. In brief, paradigm cases of confabulation involve a form of error

amounting to failure to detect object-level unreliability (motivating condition CR-m) resulting from malfunction of meta-level monitoring processes (motivating condition R-m). The rationale for the inclusion of condition V-m parallels the rationale for the inclusion of V. In brief, satisfaction of CR-m and R-m does not suffice to rule out metamnemonic luck: condition V-m is required to ensure that the subject's judgement that his episodic construction system functioned reliably must be not only accurate and adroit but also apt. Successful metamemory, in short, is to be understood as apt metamemory.

Conditions CR-m, R-m, and V-m require additional work. In particular, subjects do not normally explicitly judge, when they remember, that their episodic construction systems function reliably. Metamemory monitoring often results not in explicit metacognitive judgements but rather in metacognitive feelings (Arango-Muñoz, 2011). Ultimately, versions of CR-m, R-m, and V-m that apply to metacognitive feelings will need to be defined, but this task will have to be left for another occasion.

## 4.2 The virtue-theoretic classification of memory errors

The move to VTM was suggested in part by the role of luck in the revised simulationist classification of memory errors. In turn, however, it suggests that that classification requires further revision. The revised simulationist classification takes the fact that a reliable process produces an accurate representation (i.e., that both the reliability and the accuracy conditions are satisfied) and the fact that an unreliable process produces an inaccurate representation (i.e., that neither the reliability nor the accuracy condition is satisfied) to be sufficient for the absence of luck. But this is not right: as we saw above, luck is present when a reliable process produces an accurate representation but the representation is not accurate because the process is reliable (as in lucky LITM cases). By the same token, luck is present when an unreliable process produces an inaccurate representation but the representation is not inaccurate because the process is unreliable. The classification must thus be revised in order to take this form of luck—in addition to the form of luck that is present when a reliable process produces an inaccurate representation or an unreliable process produces an accurate representation—into account.

A revised classification is depicted in Table 3. This virtue-theoretic classification accommodates the form of luck in question by including, in addition to the accuracy and reliability conditions, an “accuracy because reliability” condition (applicable when the accuracy and reliability conditions are satisfied) and an “inaccuracy because unreliability” condition (applicable when they are not). While lucky LITM cases provide clear instances in which the “accuracy because reliability” condition is not satisfied, it is not immediately obvious what sort of case might illustrate the nonsatisfaction of the “inaccuracy because unreliability” condition; this will be left as a question for future research.

Note that there may be an important difference between lucky LITM cases and what we might intuitively think of as “lucky misinformation effect” cases. In standard misinformation effect cases (Loftus, 1996, 2005), inaccurate post-event information is incorporated into the subject's memory for an event, resulting in retrieval of

**Table 3** The virtue-theoretic classification. Light grey cells indicate object-level luck. Mid-grey cells indicate meta-level luck. Dark grey cells indicate both object-level and meta-level luck

			object level						
			R			~R			
			A		~A	A		~A	
			A b/c R	~(A b/c R)			~(~A b/c ~R)	~A b/c ~R	
meta level	R	A	A b/c R	successful remembering	e.g., lucky lost in the mall	mis-remembering			
			~(A b/c R)						
		~A							
	~R	A							
			~A b/c ~R						
			~(~A b/c ~R)				veridical confabulating		falsidical confabulating

an inaccurate memory. In a lucky misinformation effect case, accurate post-event information is incorporated into the subject’s memory for an event, resulting in retrieval of an accurate memory. Michaelian (2013) argues that, because most testimony is accurate, most post-event information received by subjects is accurate, so that the incorporation of post-event information typically results in retrieval of an accurate memory. If this is right, then, while lucky misinformation effect cases that occur in the laboratory will have a structure similar to that of lucky LITM cases, lucky misinformation effect cases that occur outside of the laboratory will not in fact involve luck—they will simply be cases in which a reliable process produces an accurate representation and in which the representation is accurate because it was produced by a reliable process, and hence they will (on either VTM or STM) be cases of successful remembering. (Standard misinformation effect cases will be cases of misremembering.)

### 5 Conclusion: a virtue theory of imagination?

Beginning with the distinction between imagination-as-failure and imagination-as-process, we arrived at VTM, an account of successful remembering as a form of imagination-as-process, and an associated classification of forms of unsuccessful remembering or imagination-as-failure. Future work by causalists might consider whether causalism can ground an analogous classification. Future work by simulationists might consider whether VTM might be generalized to provide a theory of other forms of imagination-as-process.

Because there can never be an appropriate causal connection between a subject’s representation of an event from his personal future and the event itself, the causalist approach is ill-equipped to make sense of the possibility of future-oriented confabulating (e.g., Dalla Barba, 2002, 2009, 2016) and unsuccessful episodic future

thinking more generally. Because the process that produces a subject's representation of an event from his personal future can be reliable or unreliable, the simulationist approach is better-equipped to make sense of this possibility. This virtue of the simulationist approach is, of course, inherited by the virtue-theoretic or areteist approach. Indeed, it should, in theory, be possible to generalize VTM so as to produce *a virtue theory of episodic imagination*, where episodic imagination includes not only episodic memory and episodic future thought but also episodic counterfactual thought (De Brigard, 2014a) and potentially other forms of episodic thought. In practice, doing so will require explaining how the notions of accuracy and reliability can be applied not only to episodic memory, which aims at actual past events, but also to forms of episodic thought that aim at future or counterfactual events. Areteists about memory who would be areteists about imagination thus have considerable work to do.

**Acknowledgements** Thanks for feedback to audiences at the Santa Maria-Grenoble Memory Workshop at the Universidade Federal de Santa Maria, the retreat of the Ruhr-Universität Bochum research unit FOR 2812 (Constructing scenarios of the past) in Billerbeck, the Exploring the Mind's Eye: Philosophical and Scientific Perspectives on Imagination conference at Bilkent University (especially Nazim Keven in his role as commentator), the Cologne-Grenoble Philosophy of Memory Workshop at the University of Cologne, a seminar at the University of Salzburg, the Memory, Imagination, and the Self workshop at the University of Stuttgart, and the Centre for Philosophy of Memory's internal seminar, to the students in Tom Senor's philosophy of memory seminar, to Tom Senor and Changsheng (Lex) Lai for written comments, and to two anonymous referees for exceptionally detailed reports. This work is supported by the French National Research Agency in the framework of the "Investissements d'avenir" program (ANR-15-IDEX-02) and by CAPES-COFEUCB (Grant Sh 967/20).

## References

- Addis, D. R. (2018). Are episodic memories special? On the sameness of remembered and imagined event simulation. *Journal of the Royal Society of New Zealand*, 48(2–3), 64–88.
- Addis, D. R. (2020). Mental time travel? A neurocognitive model of event simulation. *Review of Philosophy and Psychology*, 11(2), 233–259.
- Andonovski, N. (2019). Is the simulation theory of memory about simulation? *Voluntas: Revista Internacional de Filosofia*, 10(3), 37–52.
- Arango-Muñoz, S. (2011). Two levels of metacognition. *Philosophia*, 39(1), 71–82.
- Aranyosi, I. (2020). Mental time travel and disjunctivism. *Review of Philosophy and Psychology*, 11(2), 367–384.
- Battaly, H. (2008). Virtue epistemology. *Philosophy Compass*, 3(4), 639–663.
- Baysan, U. (2018). Memory, confabulation, and epistemic failure. *Logos & Episteme*, 9(4), 369–378.
- Bernecker, S. (2008). *The metaphysics of memory*. New York: Springer.
- Bernecker, S. (2010). *Memory: A philosophical study*. Oxford: Oxford University Press.
- Bernecker, S. (2017). A causal theory of mnemonic confabulation. *Frontiers in Psychology*, 8, 1207.
- Bernecker, S. (Forthcoming). An explanationist model of (false) memory. In A. Sant'Anna, C. McCarroll, & K. Michaelian (Eds.), *Current controversies in philosophy of memory*. New York: Routledge.
- Berrios, G. E. (1998). Confabulations: A conceptual history. *Journal of the History of the Neurosciences*, 7(3), 225–241.
- Cheng, S., & Werning, M. (2016). What is episodic memory if it is a natural kind? *Synthese*, 193(5), 1345–1385.
- Dalla Barba, G. (2002). *Memory, consciousness, and temporality*. New York: Springer.

- Dalla Barba, G. (2009). Temporal consciousness and confabulation: Escape from unconscious explanatory idols. In W. Hirstein (Ed.), *Confabulation: Views from neuroscience, psychiatry, psychology and philosophy*. (pp. 223–260). Oxford: Oxford University Press.
- Dalla Barba, G. (2016). Temporal consciousness and confabulation: When mental time travel takes the wrong track. In K. Michaelian, S. B. Klein, & K. K. Szpunar (Eds.), *Seeing the future: Theoretical perspectives on future-oriented mental time travel*. (pp. 119–134). Oxford: Oxford University Press.
- De Brigard, F. (2014a). Is memory for remembering? Recollection as a form of episodic hypothetical thinking. *Synthese*, 191(2), 155–185.
- De Brigard, F. (2014b). The nature of memory traces. *Philosophy Compass*, 9(6), 402–414.
- De Brigard, F. (2017). Memory and imagination. In S. Bernecker & K. Michaelian (Eds.), *The Routledge handbook of philosophy of memory*. (pp. 127–140). London: Routledge.
- De Brigard, F. (2020). The explanatory indispensability of memory traces. *The Harvard Review of Philosophy*, 27, 7–21.
- Debus, D. (2010). Accounting for epistemic relevance: A new problem for the causal theory of memory. *American Philosophical Quarterly*, 47(1), 17–29.
- Debus, D. (2017). Memory causation. In S. Bernecker & K. Michaelian (Eds.), *The Routledge handbook of philosophy of memory*. (pp. 63–75). London: Routledge.
- Dokic, J. (2014). Feeling the past: A two-tiered account of episodic memory. *Review of Philosophy and Psychology*, 5(3), 413–426.
- Fernández, J. (2018). The functional character of memory. In K. Michaelian, D. Debus, & D. Perrin (Eds.), *New directions in the philosophy of memory*. (pp. 52–71). London: Routledge.
- Fernández, J. (2019). *Memory: A self-referential account*. Oxford: Oxford University Press.
- Fernández, J. (2020). Self-referential memory and mental time travel. *Review of Philosophy and Psychology*, 11(2), 283–300.
- French, L., Garry, M., & Loftus, E. (2009). False memories: A kind of confabulation in non-clinical subjects. In W. Hirstein (Ed.), *Confabulation: Views from neuroscience, psychiatry, psychology, and philosophy*. (pp. 33–66). Oxford: Oxford University Press.
- Gallo, D. A. (2010). False memories and fantastic beliefs: 15 years of the DRM illusion. *Memory & Cognition*, 38(7), 833–848.
- Goldman, A. I. (1967). A causal theory of knowing. *The Journal of Philosophy*, 64(12), 357–372.
- Goldman, A. I. (1979). What is justified belief? In G. S. Pappas (Ed.), *Justification and knowledge*. (pp. 1–23). New York: Springer.
- Goldman, A. I. (2012). *Reliabilism and contemporary epistemology: Essays*. Oxford: Oxford University Press.
- Goodwin, D. M. (1989). *A dictionary of neuropsychology*. New York: Springer.
- Hales, S. D. (2020). *The myth of luck: Philosophy, fate, and fortune*. London: Bloomsbury Publishing.
- Hirstein, W. (2005). *Brain fiction: Self-deception and the riddle of confabulation*. Cambridge: MIT Press.
- Hirstein, W. (2011). Confabulations about personal memories, normal and abnormal. In S. Nalbantian, P. M. Matthews, & J. L. McClelland (Eds.), *The memory process: Neuroscientific and humanistic perspectives*. (pp. 217–232). Cambridge: MIT Press.
- Hutto, D. D., & Myin, E. (2017). *Evolving enactivism: Basic minds meet content*. Cambridge: MIT Press.
- Hutto, D. D., & Peeters, A. (2018). The roots of remembering: Radically enactive recollecting. In K. Michaelian, D. Debus, & D. Perrin (Eds.), *New directions in the philosophy of memory*. (pp. 97–118). London: Routledge.
- James, S. (2017). Epistemic and non-epistemic theories of remembering. *Pacific Philosophical Quarterly*, 98(S1), 109–127.
- Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin*, 114(1), 3–28.
- Klein, S. B. (2015). What memory is. *Wiley Interdisciplinary Reviews: Cognitive Science*, 6(1), 1–38.
- Korsakoff, S. (1889/1955). Psychic disturbance in conjunction with peripheral neuritis. *Neurology*, 5, 394–406.
- Liao, S., & Gendler, T. (2019). Imagination. In *Stanford Encyclopedia of Philosophy*. Centre for the Study of Language and Information. <https://plato.stanford.edu/entries/imagination/>. Accessed 14 June 2020.
- Loftus, E. F. (1996). *Eyewitness testimony*. (2nd ed.). Harvard: Harvard University Press.
- Loftus, E. F. (2005). Planting misinformation in the human mind: A 30-year investigation of the malleability of memory. *Learning & Memory*, 12(4), 361–366.

- Loftus, E. F., & Pickrell, J. E. (1995). The formation of false memories. *Psychiatric Annals*, 25(12), 720–725.
- Malcolm, N. (1977). *Memory and Mind*. Cornell University Press.
- Martin, C. B., & Deutscher, M. (1966). Remembering. *The Philosophical Review*, 75(2), 161–196.
- McCarroll, C. J. (2018). *Remembering from the Outside: Personal memory and the perspectival mind*. Oxford: Oxford University Press.
- Mercer, B., Wapner, W., Gardner, H., & Benson, D. F. (1977). A study of confabulation. *Archives of Neurology*, 34(7), 429–433.
- Michaelian, K. (2011). Generative memory. *Philosophical Psychology*, 24(3), 323–342.
- Michaelian, K. (2012). Metacognition and endorsement. *Mind & Language*, 27(3), 284–307.
- Michaelian, K. (2013). The information effect: Constructive memory, testimony, and epistemic luck. *Synthese*, 190(12), 2429–2456.
- Michaelian, K. (2016a). Against discontinuism: Mental time travel and our knowledge of past and future events. In K. Michaelian, S. B. Klein, & K. K. Szpunar (Eds.), *Seeing the future: theoretical perspectives on future-oriented mental time travel*. (pp. 63–92). Oxford: Oxford University Press.
- Michaelian, K. (2016b). Confabulating, misremembering, relearning: The simulation theory of memory and unsuccessful remembering. *Frontiers in Psychology*, 7, 1857.
- Michaelian, K. (2016c). *Mental time travel: Episodic memory and our knowledge of the personal past*. Cambridge: MIT Press.
- Michaelian, K. (2020). Confabulating as unreliable imagining: In Defence of the simulationist account of unsuccessful remembering. *Topoi*, 39(1), 133–148.
- Michaelian, K. (Forthcoming). Towards a virtue-theoretic account of confabulation. In A. Sant’Anna, C. McCarroll, and K. Michaelian (Eds.), *Current Controversies in Philosophy of Memory*. New York: Routledge.
- Michaelian, K., Perrin, D., & Sant’Anna, A. (2020). Continuities and discontinuities between imagination and memory: The view from philosophy. In A. Abraham (Ed.), *The Cambridge Handbook of Imagination*. (pp. 293–310). Cambridge: Cambridge University Press.
- 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). London: Routledge.
- Michaelian, K., & Sant’Anna, A. (Online ahead of print). Memory without content? Radical enactivism and (post)causal theories of memory. *Synthese*.
- Michaelian, K., & Sant’Anna, A. (Manuscript.) From authenticity to alethism: Against McCarroll on observer memory.
- Moscovitch, M. (1995). Confabulation. In D. L. Schacter (Ed.), *Memory distortions*. (pp. 226–251). Harvard: Harvard University Press.
- Munro, D. (Online ahead of print.) Remembering the past and imagining the actual. *Review of Philosophy and Psychology*.
- Odegard, T., & Lampinen, J. (2004). Memory conjunction errors for autobiographical events: More than just familiarity. *Memory*, 12(3), 288–300.
- Perrin, D. (2016). Asymmetries in subjective time. In K. Michaelian, S. B. Klein, & K. K. Szpunar (Eds.), *Seeing the future: Theoretical perspectives on future-oriented mental time travel*. (pp. 39–61). Oxford: Oxford University Press.
- Perrin, D. (2018). A case for procedural causality in episodic recollection. In K. Michaelian, D. Debus, & D. Perrin (Eds.), *New directions in the philosophy of memory*. (pp. 33–51). London: Routledge.
- Perrin, D., & Michaelian, K. (2017). Memory as mental time travel. In S. Bernecker & K. Michaelian (Eds.), *The Routledge handbook of philosophy of memory*. (pp. 228–239). London: Routledge.
- Pillemer, D. B., Steiner, K. L., Kuwabara, K. J., Thomsen, D. K., & Svob, C. (2015). Vicarious memories. *Consciousness and Cognition*, 36, 233–245.
- Pritchard, D. (2012). Anti-luck virtue epistemology. *The Journal of Philosophy*, 109(3), 247–279.
- Pritchard, D., & Smith, M. (2004). The psychology and philosophy of luck. *New Ideas in Psychology*, 22(1), 1–28.
- Reintz, M. T., & Lammers, W. J. (1992). Memory-conjunction errors: Miscombination of stored stimulus features can produce illusions of memory. *Memory & Cognition*, 20(1), 1–11.
- Robins, S. K. (2016a). Misremembering. *Philosophical Psychology*, 29(3), 432–447.
- Robins, S. K. (2016b). Representing the past: Memory traces and the causal theory of memory. *Philosophical Studies*, 173(11), 2993–3013.

- Robins, S. K. (2017a). Contiguity and the causal theory of memory. *Canadian Journal of Philosophy*, 47(1), 1–19.
- Robins, S. K. (2017b). Memory traces. In S. Bernecker & K. Michaelian (Eds.), *The Routledge handbook of philosophy of memory*. (pp. 76–87). London: Routledge.
- Robins, S. K. (2019). Confabulation and constructive memory. *Synthese*, 196(6), 2135–2151.
- Robins, S. K. (2020a). Defending discontinuism, naturally. *Review of Philosophy and Psychology*, 11(2), 469–486.
- Robins, S. K. (2020b). Mnemonic confabulation. *Topoi*, 39(1), 121–132.
- Schnider, A. (2018). *The confabulating mind: How the brain creates reality*. (2nd ed.). Oxford: Oxford University Press.
- Schwartz, A. (2020). Simulationism and the function(s) of episodic memory. *Review of Philosophy and Psychology*, 11(2), 487–505.
- Shanton, K., & Goldman, A. I. (2010). Simulation theory. *Wiley Interdisciplinary Reviews: Cognitive Science*, 1(4), 527–538.
- Sosa, E. (2007). *Apt belief and reflective knowledge. Vol I: A virtue epistemology*. Oxford: Oxford University Press.
- Sosa, E. (2009). *Apt Belief and Reflective Knowledge. Vol II: Reflective Knowledge*. Oxford University Press.
- Sutton, J. (1998). *Philosophy and memory traces: Descartes to connectionism*. Cambridge: Cambridge University Press.
- Turri, J. (2015). Unreliable knowledge. *Philosophy and Phenomenological Research*, 90(3), 529–545.
- Turri, J. (2016). A new paradigm for epistemology: From reliabilism to abilism. *Ergo*, 3, 189–231.
- 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.
- Zagzebski, L. (1994). The inescapability of Gettier problems. *The Philosophical Quarterly*, 44(174), 65–73.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.