Beyond the causal theory? Fifty years after Martin and Deutscher[1]

Kourken Michaelian, University of Otago

Sarah Robins, University of Kansas

Abstract: It is natural to think of remembering in terms of causation: I can recall a recent dinner with a friend *because* I experienced that dinner. Some fifty years ago, Martin and Deutscher turned this basic thought into a full-fledged theory of memory, a theory that came to dominate the landscape in the philosophy of memory. Remembering, Martin and Deutscher argue, requires the existence of a specific sort of causal connection between the rememberer's original experience of an event and his later representation of that event: a causal connection sustained by a memory trace. In recent years, it has become apparent that this reference to memory traces may be out of step with memory science. Contemporary proponents of the causal theory are thus confronted with the question: is it possible to develop an empirically adequate version of the theory, or is it time to move beyond it? This chapter traces the recent history of the causal theory, showing how increased awareness of the theory's problems has led to the development of modified version of the causal theory and ultimately to the emergence of postcausal theories.

1. Introduction

It is natural to think of remembering in terms of causation: I can recall a recent dinner with a friend *because* I experienced that dinner. Some fifty years ago, Martin and Deutscher (1966) turned this basic thought into a full-fledged theory of memory, a theory which—due both to its intuitive plausibility and its apparent success in distinguishing remembering from related processes, including imagining—came over the following decades to dominate the landscape in the philosophy of memory. Previous approaches, such as the empiricist theory,[2] had attempted to capture the nature

1 Thanks for feedback to audiences at the Université Grenoble Alpes, the University of Otago, Victoria University of Wellington, and Issues in Philosophy of Memory (Cologne 2017), and thanks for written comments to Steven James and Denis Perrin.

2 For background on the empiricist theory, see Bernecker, 2008. Bernecker also discusses the epistemic theory, which has likewise been eclipsed in popularity by the causal theory.; the epistemic

of remembering from a first-person perspective, in terms of its characteristic phenomenology. The causal theory, in contrast, offered a third-personal account of the nature of remembering. Remembering, Martin and Deutscher argue, boils down to the existence of a specific sort of causal connection between the rememberer's original experience of an event and his later representation of that event: a causal connection sustained by a memory trace.

Though it initially seemed to boost the theory's fit with the empirical science of memory, it has become apparent in recent years that this reference to memory traces in fact threatens to undermine the causal theory. As older conceptions of memory in terms of storage and retrieval have given way to new conceptions of remembering as a constructive or simulational process, contemporary memory science has appeared to overturn not only the particular view of traces advocated by Martin and Deutscher but also the more general claim that traces—of one sort or another—are essential to remembering. Contemporary proponents of the causal theory have thus been confronted with the question: is it possible to develop an empirically adequate version of the theory, or is it time to move beyond the causal theory? The purpose of this chapter is to trace the recent history of the causal theory, showing how increased awareness of problems for the classical causal theory has led to the development of a variety of updated versions of the theory and ultimately to the emergence of postcausal theories.

2. The classical causal theory

Like most subsequent causal theorists, Martin and Deutscher focus on episodic memory, memory for past events. Omitting certain technical details, they argue that a subject remembers a past event if and only if (1) he now represents the event, (2) he experienced the event when it took place, and (3) there is a causal connection between his current representation of the event and his experience of it. This account treats memory as a diachronic capacity, in the sense that it claims that for

---

theory is not to be confused with the hybrid causal-epistemic theory reviewed in section 5 below. Martin and Deutscher were not the first to state the causal theory, but they offer the canonical statement of the theory, and so we do not discuss earlier formulations here.

remembering to occur is for there to be a particular relationship between representations located at two different points in time: the subject's original experiential representation of the event and his current retrieved representation[3] of the event. Conditions (1) and (2) require the existence of these representations and are widely accepted constraints on remembering. It is in virtue of the third condition, stipulating a causal connection between the two representations, that Martin and Deutscher's account qualifies as a *causal* theory. Anticausal approaches to the mind in general (e.g., Wittgenstein, 1953; Holland, 1954) and to memory in particular (e.g., Malcolm, 1963; Squires, 1969) were popular when Martin and Deutscher wrote, and the causal condition was therefore objectionable to many of their contemporaries. Nevertheless, though acausal accounts of memory are still occasionally defended (e.g., Martin, 2001; Hamilton, 2003), the causal theory, arguably due to the attention devoted by Martin and Deutscher to refining condition (3), gradually won out over the alternatives.

In formulating the causal condition, Martin and Deutscher's primary concern was to differentiate *remembering* from *imagining*. Even if a subject somehow manages to produce a representation that is accurate with respect to a past experience that she has had, her representation will intuitively fail to qualify as a memory if it lacks a causal connection to that experience. Suppose that Roger attends a magic show. Later, he suffers an accident, the result of which is complete retrograde amnesia: he no longer remembers events from his past, including the magic show. Also as a result of the accident, he is prone to producing confabulatory accounts of past events. Suppose that he produces a story that happens to correspond in perfect detail to his experience of the magic show. Conditions (1) and (2) are satisfied, but Roger is clearly not remembering. This sort of coincidental correspondence between an experiential representation and a retrieved representation may be unlikely, but its very possibility suggests the need for a causal

3 Throughout, "retrieved representation" refers to the representation entertained by the subject at the time of (apparent) remembering, regardless of whether the process responsible for the production of the representation in question in fact involved retrieval of information and regardless of whether remembering in general is understood as involving retrieval.

connection between the representations—absent such a connection, the subject would seem to be merely imagining. Hence condition (3).

Martin and Deutscher argue further that not just any causal connection between an experiential representation and a retrieved representation suffices for remembering: remembering requires a causal connection sustained by a *memory trace*. The inclusion of a reference to memory traces in the theory is necessary in part in order to differentiate remembering from *relearning*, which occurs when one acquires information through experience, forgets it, and then reacquires it from another source. Suppose, again, that Roger attends a magic show; later, he suffers an accident, the result of which is complete retrograde amnesia. If, at some point between the show and the accident, Roger told his friend Lane about the show, then he might later relearn of it from him. Suppose that Lane comforts Roger after his trauma by repeating stories of his past, including that of the magic show. As a result, Roger is again able to represent the event. In this case, the experiential representation and the retrieved representation are causally connected: Roger's experience of the magic show is the cause of his conversation with Lane, which is in turn causally implicated in Lane's relaying the information to him. But intuitively this is a case of relearning, not remembering.

In differentiating remembering from relearning, Martin and Deutscher were sensitive to the fact that the occurrence of remembering is compatible with the use of external prompts. Drawing a distinction between remembering and relearning requires saying when external information serves as a mere supplement to memory and when it serves as a replacement for it; that is, we need a way of excluding *relearning* while permitting *prompting*. Martin and Deutscher do not draw the distinction in terms of the quantity of external information involved in the process of (apparent) remembering but rather in terms of the role it plays. Remembering, for them, is compatible with extensive prompting from external sources. What matters is whether there is also an internal state of the (apparent) rememberer that is active—a state acquired as a result of the experience that he is trying to remember, that is, a memory trace. Condition (3) thus becomes: there is a causal

connection, sustained by a memory trace, between the subject's retrieved representation of the event and his experiential representation of the event.

Just as the bare requirement of a causal connection was objectionable to many of Martin and Deutscher's contemporaries, so was the more specific requirement of a causal connection sustained by a memory trace. Some worried that to include a reference to memory traces in a philosophical theory of remembering was to allow philosophy to "dictate to science what to discover in the human brain" (Zemach, 1983: 32). Others were concerned about influence in the opposite direction, worrying that Martin and Deutscher's reference to memory traces was an attempt to import a scientific notion into the everyday concept of memory that philosophy was meant to analyze (Malcolm, 1977). The relationship of the causal theory of memory to the science of memory remains an open question, and we return to this question in subsequent sections.

Martin and Deutscher were also sensitive to the possibility that a cognitive capacity other than memory, also acquired during the subject's experience of an event, might result in a later representation of the event. The desire to preclude this possibility led them to add further details to the memory trace requirement. Suppose that Roger, while attending the magic show, is hypnotized and as a result can be placed in a highly suggestible state. Suppose that Lane tells Roger about the magic show while he is in this suggestible state and that Roger endorses Lane's account. Intuitively, though there is a causal connection between his experience of the magic show and his representation of it, he does not remember the magic show. This case fails to be a case of remembering because, while Roger might have a suitable memory trace, his memory trace is not doing the relevant causal work—it is some other, *nonmemorial* capacity that is responsible for the representation. To exclude such cases of nonmemorial retention,[4] Martin and Deutscher argue that remembering requires the preservation of a trace that *represents* the past and provides the content of the retrieved representation. In particular, they see traces as "structural analogues" of past

4 For an extended discussion of nonmemorial retention, see Robins, 2016b.

experiences: a memory trace is an entity that contains a quantity of information that matches or exceeds what the subject recalls about the relevant event. In other words, remembering, for them, necessarily involves the transmission of content from experience to retrieval and is incompatible with the generation of new content between experience and retrieval.

Martin and Deutscher's appeal to memory traces is simultaneously a nod to convention and a bold innovation. On the one hand, the claim that memory traces are structural analogues of past experience is a longstanding and widespread assumption of both philosophical and everyday thinking about memory (see Draaisma, 2000; De Brigard, 2014b): just as Martin and Deutscher compare memory to the grooves of a record, Plato, for example, compared it to impressions in a wax tablet. On the other hand, Martin and Deutscher offer a new reason for this old view of memory traces, treating traces not as the *objects* of remembering but rather as the *bearers* of the right sort of causal connection between the experiential representation and the retrieved representation. Despite the fact that the characterization of memory traces as structural analogues of past experiences is traditional, however, there is reason to prune it from the causal theory. To say that memory traces are structural analogues of past experiences is to say that a memory trace represents an experience in virtue of its standing in a relationship of structural isomorphism with that event. As an account of mental representation, structural isomorphism provides a way of ensuring that the inferential interactions between the contents of thought are reflected in the causal interactions between the vehicles by which they are represented. Although this view of mental representations was popular at the time at which Martin and Deutscher were writing, it is controversial and is not now widely endorsed (e.g., Shepard & Chipman, 1970). Moreover, the characterization of memory traces as structural analogues of past experience makes a claim about how mental representation works, and this specific claim goes beyond the general claim, required by the causal theory, that memory traces must be mental representations.[5] In what follows, we

[5] For a detailed argument to this effect, see Robins, 2016a.

therefore do not interpret the classical causal theory as incorporating a characterization of memory traces as structural analogues of past experiences.

According to the classical causal theory, then, it is memory traces that make the difference between a mere causal connection between an experiential representation and a retrieved representation and what we can refer to as an *appropriate* causal connection, a causal connection of the sort required to underwrite remembering. Pruned of the structural analogy requirement, the causal theory makes an empirical bet regarding the existence of traces but stops short of betting on any particular account of the physical nature of memory traces. The physical details do not matter; only certain very general features do. In line with their treatment of imagining, relearning, and nonmemorial retention, Martin and Deutscher are committed to viewing memory traces, first, as being distinct *states* and, second, as having distinct *contents*.

Regarding the first commitment, a memory trace must be a distinct, internal state of the rememberer. In order for this condition to be met, the causal chain leading back to the experience must be distinguishable from other causal chains. After all, people have multiple memories and therefore multiple memory traces. Roger, from our example above, has a memory of attending a magic show but presumably many other memories as well. To determine whether he remembers the magic show requires establishing that that this particular causal chain has been sustained. To determine whether he remembers another experience—his 5th birthday, for instance—requires establishing the existence of a different causal chain. This is only possible if the chain supported by each internal state is distinct. This distinctness serves as a marker of the unique causal history of each memory trace, which becomes especially important in establishing the difference between remembering and relearning. Remembering and relearning might produce exactly similar representations; the only way of differentiating between them is by when and how they were acquired.

Regarding the second commitment, the memory trace must not only provide a distinct causal

link via an internal state that serves as a representation of that experience. As in the earlier example of hypnosis, it is possible that other aspects of a preserved internal state could result in a representation of a past experience. If we are to establish the difference between remembering and nonmemorial forms of retention, then there must be some way in which the memory trace is distinct from these other processes. The memory trace must be a distinct component of the internal state in which it features, distinguishable from all other components this state may have. The memory trace is distinctive, Martin and Deutscher argue, because it alone represents that past experience. By preserving information about that event or experience across time, the memory trace is distinguishable from other retained states that could in one or another way result in representations of the experience. Moreover, by preserving information over time, the memory trace provides an explanation of how accurate retention of information from the past is possible.

3. Neoclassical causal theories

Fifty years after Martin and Deutscher wrote, the classical causal theory continues to be influential, and a number of causal theories that may be characterized as *neoclassical* have recently been developed. Neoclassical causal theories retain the core claim of Martin and Deutscher's theory—that an appropriate causal connection (where appropriate causation is understood as causation going via a memory trace) is both necessary and, along with other suitable conditions, sufficient for memory—while modifying certain less central elements of the theory. The theories proposed by Bernecker (2008, 2010) and Cheng and Werning (2016) are illustrative of the neoclassical approach.

Offering a systematic argument for the superiority of the causal theory over noncausal theories, Bernecker offers a detailed development of a causal theory in the spirit of Martin and Deutscher's. In particular, he understands appropriate causation in terms of contiguity, maintaining that it is the presence of an uninterrupted chain of memory traces between learning and remembering that distinguishes remembering from imagining and relearning. Bernecker's analysis also updates Martin and Deutscher's in certain respects. First, he denies that the content of the

experiential representation and the content of the retrieved representation must be identical. Instead, they must be "sufficiently similar" (2010: 217): content can change over time (e.g., one might initially remember receiving a new bicycle and later only remember receiving a gift), but no new content can be generated. Second, he endorses a distributed view of traces. Unlike the distributed conceptions of traces that we discuss in the next section, however, Bernecker's view is that traces are distributed at the implementational level only, allowing content transmission to occur at the psychological level.

Cheng and Werning's approach differs from Bernecker's in terms of both scope and method. In terms of scope, Bernecker discusses a range of forms of memory, including memory for persons and things, memory for properties, memory for events, and memory for facts and propositions, focusing on the latter. Cheng and Werning focus specifically on memory for events—more precisely, on episodic memory, their understanding of which, in line with the psychological literature on mental time travel (Suddendorf & Corballis, 1997), includes a role for autonoesis, or consciousness of the self in subjective time (Tulving, 1985), a topic to which we return in section 3. In terms of method, whereas Bernecker relies primarily on the tools of conceptual analysis. Cheng and Werning's approach is naturalistic in spirit, appealing to data on the role of specific brain structures, primarily the hippocampus, in remembering; like Michaelian (2011b), they seek to understand memory as a natural kind. While this naturalistic approach lends a degree of methodological novelty to their approach, the main substantive novelty of their version of the causal theory consists in its characterization of memory representations as being sequential in nature, a characterization which they derive from their understanding of the role of hippocampal processes in remembering (cf. Cheng et al., 2016). Ultimately, however, the gist of their theory—which requires that the retrieved representation be causally grounded in the corresponding earlier experience via a memory trace—is similar to that of Bernecker's, which, as we have seen, is in turn similar to that of Martin and Deutscher's.

Both Bernecker (2008, 2010) and Cheng and Werning (2016) can thus be classified as neoclassical causal theorists,[6] and the differences between their respective versions of the causal theory, as well as those between their versions of the causal theory and Martin and Deutscher's version of the theory, can be set aside for present purposes. Both classical and neoclassical causal theorists assume, first, that remembering involves the *transmission* of content from experience to retrieval and, second, that remembering is incompatible with the *generation* of new content between experience and retrieval. Each of these assumptions has, however, been denied by other recent versions of the causal theory. We consider theories that deny the former assumption in section 5 and theories that deny the latter in section 6.

4. Hybrid theories

Setting the issues of transmission and generation aside for the moment, we emphasize that Martin and Deutscher's core claim—that an appropriate causal connection is both necessary and, along with other suitable conditions, sufficient for memory—is accepted in one form or another by many contemporary philosophers of memory (see Debus, 2017).[7] In particular, we note that the literature contains few challenges to the claim that appropriate causal connection is *necessary* for memory. Some invoke this claim in passing while focusing on other issues (e.g., Debus, 2008, 2014; Hopkins, 2014). Others do not invoke it but nevertheless say nothing to challenge it. In contrast, the literature does contain a number of challenges to the claim that appropriate causation is *sufficient* for memory. If one of these challenges were to succeed, it would be necessary to supplement the appropriate causation condition—along with the other basic conditions required by the causal theory—with a further condition, thus producing a *hybrid* theory of remembering.

Debus (2010; cf. James, forthcoming), for example, argues that genuine memories are, in addition to being causally connected to the subject's past experiences, necessarily *epistemically*

6 Cf. Deutscher's (2017) comparison of Bernecker's to Martin and Deutscher's theory, which provides a more detailed discussion of the points of similarity between the two.
7 Some have argued for a return to epistemic (e.g., Adams, 2011) or even empiricist theories (Byrne, 2010) of remembering, but such arguments are infrequent.

relevant to the subject, in the sense that he is disposed to take them into account when forming judgements about the past, typically (but not always) by forming a belief that the remembered event occurred. Because the classical causal theory does not treat epistemic relevance as necessary for remembering, Debus maintains, it is bound to classify certain cases as instances of genuine memory when in fact they are instances of merely apparent memory. (Consider Martin and Deutscher's oft-discussed case of a painter who paints a scene from his past without realizing that it is a scene from his past.) This argument, which, if it works, applies equally to neoclassical causal theories, in effect suggests that the *causal* theory should be replaced with a hybrid *causal-epistemic* theory.

Similarly, Klein (2014, 2015; cf. Dokic, 2014) argues that genuine memories necessarily involve, in addition to causal connection, a specific phenomenology: *autonoetic consciousness*, or a sense of the self in subjective time. Klein and Nichols (2012; cf. Fernández, forthcoming), for example, discuss the case of patient RB, whom they characterize as having retained the capacity to retrieve information deriving from his past experiences but as lacking a "sense of mineness" for the memories thus produced. Though the causal theory would classify the case of RB as one in which the subject is capable of remembering, on Klein's view RB is, because he lacks the capacity for autonoesis, incapable of genuine memory. This argument, which, if it works, applies, like Debus's argument, equally to the other versions of the causal theory considered so far, in effect suggests that the causal theory should be replaced with a hybrid *causal-autonoetic* theory.

The causal-autonoetic theory and the causal-epistemic theory are close cousins: as Mahr and Csibra (forthcoming) have emphasized, the involvement of autonoesis in remembering explains the subject's tendency to believe that remembered events occurred. And they are thus vulnerable to similar challenges. In particular, both the causal-autonoetic and the causal-epistemic theory imply that semantic memory (memory for facts) differs radically in kind from episodic memory (memory for experienced events) (Michaelian, 2015), suggesting that they conflate a requirement for *episodicity* (Perrin & Rousset, 2014) with a requirement for *mnemicity* (Michaelian & Sutton, 2017).

Even if autonoesis or epistemic relevance turns out to be a requirement for mnemicity, however, there would appear to be nothing that would prevent an advocate of a given version of the causal theory (or a given postcausal theory; see below) from adding an appropriate condition to his theory. Hybrid theories will therefore be set aside in what follows.

5. Distributed and procedural causal theories

Though it accepts the sufficiency of appropriate causation, there is a sense in which the distributed causal theory departs more radically from the classical causal theory than do hybrid theories. Hybrid theorists posit conditions on remembering in addition to the appropriate causation condition. Distributed causal theorists take a different tack, modifying the concept of a memory trace in such a way that the appropriate causation condition can arguably no longer be understood as requiring transmission of content from experience to retrieval. We say "arguably", for distributed causal theorists have not always been clear about whether they deny that appropriate causation involves transmission of content. Indeed, the literature contains no detailed articulation of the distributed causal theory. Sutton (1998, 2010) has provided a detailed account of *the distributed conception of traces* but has said little about how this conception of traces might be combined with *the causal theory*. Bernecker (2010) and Michaelian (2011), meanwhile, have developed detailed versions of the causal theory that endorse distributed traces in principle but have said little about their own distributed conceptions of traces.

Notwithstanding this gap in the literature, it is uncontroversial that the inspiration for the distributed causal theory comes primarily from debates and developments regarding the nature of mental representation more generally. Much as a general view of mental representation in terms of structural analogy influenced Martin and Deutscher's account of traces as structural analogues of experience, proponents of distributed conceptions of traces have been influenced by connectionist, dynamicist, and distributed views of mental representation. The traditional conception of traces involves fixed, explicit contents carried by distinct local vehicles. The vehicles in question might be

distributed in the sense that they are complex entities the parts of which are stored in different locations, but they are local in the sense that each memory content is carried by a distinct vehicle. Proponents of distributed conceptions challenge this matrix of ideas, arguing that we should give up at least some of the features of the traditional conception.

Sutton's account of distributed traces comes closest to a full-blown rejection of the traditional conception: memories, he argues, "are blended, not laid down independently once and for all, and are reconstructed rather than reproduced'' (1998: 2). On this account, a subject's memory is a network in which various items of information are connected as a function of the frequency with which they co-occur in his experience. Each experience activates a certain pattern in the network, but the patterns overlap in a way that precludes distinct contents or vehicles. If this view is right, we may be able to refer to memory traces in a loose sense, since a specific experience will result in a specific modification of connections in the network, but these are traces of a sort that require us to reject the two key commitments of (neo)classical causal theories (identified in section 2): there are no traces in the sense of *distinct vehicles* carrying *distinct contents*. Due to the gap in the literature noted above, it remains unclear how, in view of the fact that they reject these commitments, distributed causal theorists would have us understand the nature of the causal connection that they take to hold between retrieved memories and experiences, and there is a pressing need for further work on this question.

Some distributed causal theorists have been less specific about the nature of memory traces but have tried to reconcile a distributed conception of traces with the appropriate causation condition. These authors reject the (neo)classical assumption that remembering involves the transmission of content from experience to retrieval, instead maintaining that content is reconstructed at the time of retrieval. To say that remembering is *reconstructive*, rather than *reproductive*, is to say that the content of a retrieved representation is, at least in part, produced at the time of retrieval, rather than transmitted from the corresponding experience. There is a long-

standing consensus in the empirical literature that remembering is reconstructive in this sense (see, e.g., Schacter & Addis, 2007; Schacter et al., 2012). One possible reaction to the reconstructive character of remembering would be to continue to understand traces as distinct entities but to hold that their content is implicit in the sense that it needs to be activated or made explicit at the time of retrieval (see Vosgerau, 2010). Another possible reaction is provided by the procedural causal theory developed by Perrin (this volume).

The procedural causal theory explicitly denies that remembering involves the transmission of content. Perrin retains a generic version of the core claim of the causal theory—that an appropriate causal connection is both necessary and, along with other suitable conditions, sufficient for memory—but understands it in a radically different manner than do (neo)classical causal theorists. Whereas (neo)classical causal theorists understand causal connection in terms of the transmission of content via memory traces, procedural causal theorists take the reconstructive character of remembering to undermine this understanding of causation in memory. Inspired by older attributionalist approaches in psychology (e.g., Kolers & Roediger, 1984; Jacoby et al., 1989; Whittlesea, 1997), Perrin proposes an alternative understanding of the nature of causation in memory. The key idea is that, rather than the *content* of the retrieved representation being causally related to the *content* of the corresponding experience, it is the *process* that produces the retrieved representation that is causally related to the *process* that produced the corresponding experience. Adopting a view of perception as itself being a constructive process, Perrin's suggestion is that the constructive process of perceiving may bear certain similarities to the reconstructive process of remembering and thus give rise to a degree of fluency in the latter—it is in general easier to reconstruct a scene that one has previously constructed—despite the fact that no content is transmitted.

The procedural causal theory may succeed in providing a description of a kind of *causal connection* that can obtain between experience and retrieval despite the fact that no content is

transmitted from the former to the latter. But it does not yet provide a description of what it is for such a causal connection to be *appropriate*. While this is a potential problem, perhaps a more pressing question for both procedural and distributed causal theorists is whether they mean to retain the (neo)classical assumption that remembering is incompatible with the generation of new content between experience and retrieval. In one sense, of course, distributed and procedural causal theories necessarily acknowledge that remembering involves the generation of content, since they claim that content from previous experience is retained at best only implicitly, which implies that content must be "regenerated" at the time of retrieval. But this is just to say that they deny what might be called *"transmissionism"*, the view that (explicit) content is stored between experience and retrieval. In another sense—and this is the sense that matters here—they may deny that remembering involves the generation of content, since it is open to them to deny that "regenerated" retrieved content may include information going beyond that of the experience. That is, it is open to them to accept *preservationism*, the view that a retrieved representation may not include content not included in the original experience.

More conservative versions of the theories will accept preservationism, but the basic distributed and procedural causal theories can be conjoined with a range of views on the generation of content. The more extreme the views become, the more likely they are to reject the core commitments of the causal theory. The most conservative view available is that the content of the retrieved representation is identical to the content of the experiential representation. This extreme form of preservationism is incompatible even with the occurrence of forgetting and is not to be taken seriously. An intermediate view is that the content of the retrieved representation must be contained in or in some sense implied by the content of the experiential representation. This more moderate form of preservationism is compatible with the occurrence of forgetting but not with the generation of new content between experience and retrieval and is explicitly endorsed by some (e.g., Bernecker 2008, 2010; Cheng & Werning, 2016) and implicitly assumed by many others. While it

is always possible in principle to hold on to preservationism by enriching the content of experience (McCarroll, 2017), we will see below that there is a real tension between even the moderate form of preservationism and the reconstructive character of remembering, which suggests a form of *generationism* according to which the content of the retrieved representation may indeed include information not included in the content of the experiential representation.[8]

6. Postcausal theories

As we saw in section 4, the sufficiency of appropriate causation is challenged by hybrid theories on phenomenological or epistemic grounds. A different sort of challenge to the sufficiency of appropriate causation arises due to the reconstructive character of remembering, i.e., due to the fact that the content of retrieved representations is, at least in part, produced at the time of retrieval, rather than derived from the content of the corresponding experience. Reconstruction, in fact, challenges not only the sufficiency of appropriate causation but also its necessity and has therefore led to the emergence of theories that may be characterized as *postcausal*, in the sense that they claim that a causal connection—"appropriate" or otherwise—is not necessary for memory, even while recognizably descending from the causal theory. Postcausal theories in effect treat memory as a synchronic rather than a diachronic capacity, in the sense that they see the occurrence of remembering as depending on what happens when the subject (apparently) remembers, rather than on whether there is a suitable relationship between the subject's retrieved representation and his experiential representation; thus, unlike hybrid theories, they move decisively beyond the causal theory.

One intriguing postcausal theory is the functionalist theory, which Fernández (this volume) offers as an alternative to both the causal theory and the narrative theory of memory (e.g., Schechtman, 1994; Goldie, 2012; Brockmeier, 2015). Fernández argues that the causal theory is

8 "Preservationism" sometimes refers to the view that memory preserves justification, as opposed to the view that it preserves content (see Lackey, 2005; Fernández, 2016; Frise, forthcoming). We are concerned here neither with this form of preservationism nor with the corresponding form of generationism.

both too strict, in that it is incompatible with the generation of new content during reconstructive remembering, and too weak, in that it ignores the tendency (emphasized by hybrid theories) for memory to give rise to belief. He likewise argues that the narrative theory—which, emphasizing reconstruction, views remembering as an imaginative process in which the subject draws on stored information deriving from his experiences, along with information deriving from other sources, to create narratives about his past—is both too strict, in that it does not acknowledge the possibility of memories that are not embedded in narratives, and too weak, in that it does not acknowledge any role at all for the causal history of memories. The alternative that Fernández offers is a theory on which a mental state qualifies as a memory just in case it plays the *functional role* that memories typically play, where this role is a matter, first, of tending to cause belief and, second, of tending to be caused by past experience. What is most important about the functionalist theory, in the present context, is the second of these claims: while the functionalist theory requires, in order for a mental state to qualify as a memory, that it *tend* to be caused by the subject's past experience of the remembered event, it does not require that the mental state *actually* be caused by the experience. The functionalist theory thus rejects the core claim of the causal theory.

In line with the discussion of the causal-epistemic theory above, the second of the functionalist's claims, regarding the link between memory and belief, may be understood as concerning episodicity, rather than mnemicity. If we therefore disregard this claim, Fernández' functionalist theory and the simulation theories recently developed by a number of authors (Shanton & Goldman, 2010; De Brigard, 2014; Michaelian, 2016) come to broadly similar conclusions about the nature of remembering. The path taken by the simulation theorist is, however, somewhat less direct, involving a close consideration of the role of traces in remembering. It might be thought, given the association between reconstruction and distributed/procedural theories, that local trace theories can avoid the challenge posed by reconstruction, but the causal theory cannot in fact be protected by retreating to the local conception. Even if, as noted above, the distributed conception

has in many cases been adopted only in a nominal sense, most philosophers of memory have been in principle convinced by the arguments in favour of the distributed conception. And even if some have not yet been convinced by the arguments and so deliberately continue to work with the local conception, they are nevertheless bound, given the weight of the evidence in its favour, to acknowledge the reconstructive character of remembering within the parameters of the local conception. The challenge must thus be faced by all causal theorists.

Is the existence of an appropriate causal connection between the retrieved representation and the experiential representation sufficient for remembering, given the local conception of traces? Given reconstruction, the local trace theorist must acknowledge what we might refer to as "the fact of multiple experiences": multiple experiences may contribute to the content of a single stored trace. He must also acknowledge what we might refer to as "the fact of multiple traces": multiple traces may contribute to the content of a single retrieved representation. These facts together imply that, if a given retrieved representation is appropriately causally connected to a given experience, it may also be appropriately causally connected to other experiences. The existence of an appropriate causal connection thus does not suffice, given the local conception, to determine whether the subject is remembering a given event.

Is the existence of an appropriate causal connection between the retrieved representation and the experiential representation sufficient for remembering, given the distributed conception of traces? Given the distributed conception, retrieval is a matter of activating certain ideas—nodes in a larger network of ideas—together. The tendency for certain ideas to be activated together is, however, not attributable to a unique event, since the relevant connection weights have inevitably been affected by multiple experiences (Robins, 2016b). Nor is there any guarantee that a given retrieved representation matches a unique experiential representation. It is, as noted in section 5 above, not entirely clear how the notion of appropriate causation is to be understood by the distributed trace theorist. But however it is understood, it would appear that the distributed conception implies that,

if a given retrieved representation is appropriately causally connected to a given experience, it may also be appropriately causally connected to other experiences. The existence of an appropriate causal connection thus does not suffice, given the distributed conception, to determine whether the subject is remembering a given event.

If appropriate causation were merely to fail to be sufficient for memory, it would be possible to save the causal theory by means of the incorporation of an additional condition, in the manner of the hybrid theories discussed in section 3. But reconstruction appears to undermine not only the sufficiency of appropriate causation but also its necessity. Beginning with the local conception of traces, the fact of multiple experiences and the fact of multiple traces together imply that the content of a retrieved representation will typically not derive entirely from that of the relevant earlier experience. In some cases, a majority of the content may so derive. In other cases, however, only a minority of the content so derives. And in some cases, none of the content so derives. As long as some of the content derives from the experience, of course, a causal connection obtains, and it is intuitively plausible that there is a difference in kind between such cases and cases in which none of the content derives from the experience. On the basis of this intuition, Michaelian (2011a) has argued for a constructive causal theory, a theory which is like the causal theory in that it requires the *transmission* of content from experience to retrieval (accepting transmissionism) but unlike it in that it permits the *generation* of new content between experience and retrieval (rejecting preservationism and accepting generationism).

Robins (2016a), who likewise seeks to retain the causal theory while acknowledging the reconstructive character of remembering, has defended a broadly similar approach. While constructive causal approaches provide an appealing means of reconciling the causal theory with reconstruction, however, the empirical research on reconstruction suggests, as Michaelian has pointed out in subsequent work (2016c) that the very same cognitive process may be at work both in cases in which *some* content is transmitted from the experience and in cases in which *no* content

is transmitted. This in turn implies that, given the local conception of traces, causal connection does not mark the difference between genuine and merely apparent memory. Turning to the distributed conception of traces, we find a similar implication. Due to the blended nature of distributed storage, not all of the ideas that compose a given retrieved memory are activated because of the relevant earlier experience. In some cases, a majority of the ideas may be activated due to the earlier experience. In some cases, however, only a minority are. And in some cases, none are. There is, however, no reason to suppose that there is a difference in kind between cases in which none of the ideas are activated because of the relevant earlier experience and cases in which at least some are—in cases of both sorts, the same process may be at work. This implies that, given the distributed conception of traces, causal connection does not mark the difference between genuine and merely apparent memory.

A distributed or procedural causal theorist might object that this argument presupposes transmissionism, which distributed and procedural theories reject. The idea would be that a distributed/procedural theory can reject transmission but accept either preservationism or a moderate form of generationism according to which there must be some degree of overlap between the content of the retrieved representation and the content of the earlier representation in order for genuine remembering to occur. The distributed/procedural theorist can then maintain that genuine remembering occurs only if, first, there is such overlap and, second, this overlap is due to the presence of an appropriate causal connection, understood in nontransmissionist terms. While this is an interesting objection, it assumes that a convincing nontransmissionist account of appropriate causation can be formulated, and this remains to be done. The argument given above does not presuppose transmissionism but does bet that there will turn out to be no interesting difference between cases in which the activation of at least some of the relevant ideas is due to the earlier experience and cases in which the activation of none of them is due to the earlier experience.

Reacting to these difficulties for the constructive causal theory, Michaelian has proposed a

simulation theory of remembering, the key idea of which is that, contrary to the basic assumption of the causal theorist, there is no difference between remembering the past and imagining it, in which case memory does not presuppose a causal connection—to remember just is to imagine the past. De Brigard (2014a), though he is less explicit about his stance on the necessity of causal connection, has developed a similar view, treating episodic memory as a form of episodic hypothetical thought, or thought about possible events. And Shanton and Goldman (2010) have likewise argued that remembering is to be understood in simulational terms, linking remembering to theory of mind. Evidence for the simulation theory comes from research on episodic memory as a form of mental time travel analogous to episodic future thought (Suddendorf & Corballis, 1997). A large body of research now supports the view that the process of remembering the past is executed by the same cognitive system as the process of imagining the future and, indeed, that imagining the future is the primary function of the system in question (see Michaelian et al., 2016). Both imagining the future and remembering the past draw on stored content originating in experience of past events. Imagining a future event, does not, of course, draw on content originating in experience of the particular event imagined. By the same token, the mental time travel framework suggests that remembering a past event does not necessarily draw on content originating in experience of the particular event remembered. From a broadly naturalistic point of view, this, in turn, suggests that remembering does not presuppose a causal connection.

If remembering does not presuppose a causal connection, a fortiori it does not presuppose an *appropriate* causal connection. But this does not mean that the process of imagining the past cannot itself be appropriate or inappropriate: if the subject imagines the past in the wrong way, the representation he produces may fail to qualify as a memory, even if it should happen to be accurate. Not only simulation theorists but also constructive causal theorists, who acknowledge that memories may be in part the product of imagination, even if they deny that they can be wholly the product of imagination, thus must provide an account of the appropriateness of the process of

imagining the past. Michaelian's version of the constructive causal theory therefore incorporates a reliability condition—a condition requiring that the system function in such a way that it tends to produce mostly accurate representations—and this condition is inherited by his version of the simulation theory, which, strictly speaking, says that to remember a past event is to imagine it *in a reliable manner*. The reliability condition enables the simulation theory to distinguish remembering, understood as imagining the past, from confabulation and other ways of *merely* imagining the past. It remains to be seen whether further conditions must be added to the simulation theory in order to enable it to distinguish between remembering and relearning and between remembering and nonmemorial retention.

[Figure 1 about here.]

7. Conclusions

Fifty years after Martin and Deutscher, causal theories of various sorts—neoclassical, hybrid, and distributed/procedural—continue to dominate the landscape in the philosophy of memory (see figure 1). Clearly, the field as a whole has yet to move decisively beyond the causal theory. The emergence of postcausal theories, however, hints at increased awareness of the tension between the causal theory and the reconstructive character of remembering. Of course, while postcausal theories may be better suited than causal theories to accommodating the reconstructive character of remembering, they will themselves inevitably face objections. The functionalist theory is too new for objections to it to have emerged. But objections to the simulation theory—focusing on the "continuist" view of past- and future-oriented mental time travel that it presupposes (Perrin, 2016; Michaelian, 2016a; Perrin & Michaelian, 2017) and on its ability to distinguish between remembering and misremembering or confabulating (Robins, 2016b; Michaelian, 2016b; Robins, forthcoming)—have already begun to be voiced. Time will tell whether postcausal theorists are able to address these and other objections and convince significant numbers of philosophers of memory to move beyond the causal theory.

References

Adams, F. (2011). Husker du? *Philosophical Studies*, *153*(1), 81-94.

Bernecker, S. (2008). *The Metaphysics of Memory*. Springer.

Bernecker, S. (2010). *Memory: A Philosophical Study*. Oxford University Press.

Brockmeier, J. (2015). *Beyond the Archive: Memory, Narrative, and the Autobiographical Process*. Oxford University Press.

Byrne, A. (2010). Recollection, perception, imagination. *Philosophical Studies*, *148*(1), 15-26.

Cheng, S., & Werning, M. (2016). What is episodic memory if it is a natural kind? *Synthese*, *193*(5), 1345-1385.

Cheng, S., Werning, M., & Suddendorf, T. (2016). Dissociating memory traces and scenario construction in mental time travel. *Neuroscience & Biobehavioral Reviews*, *60*, 82-89.

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.

Debus, D. (2008). Experiencing the past: A relational account of recollective memory. *Dialectica*, *62*(4), 405-432.

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. (2014). 'Mental time travel': Remembering the past, imagining the future, and the particularity of events. *Review of Philosophy and Psychology*, *5*(3), 333-350.

Debus, D. (2017). Memory Causation. *Routledge Handbook of Philosophy of Memory*. Eds. S. Bernecker and K. Michaelian. Routledge. Pp. 63–75.

Deutscher, M. (2017). The trace as structural analogue. *The Routledge Encyclopedia of Philosophy*. Taylor and Francis. https://www.rep.routledge.com/articles/thematic/memory/v-2/sections/the-trace-as-structural-analogue

Dokic, Jérôme. (2014). Feeling the past: a two-tiered account of episodic memory. *Review of Philosophy and Psychology*, *5*(3), 413-426.

Draaisma, D. (2000). *Metaphors of Memory: A History of Ideas about the Mind*. Cambridge University Press.

Fernández, J. (2016). Epistemic generation in memory. *Philosophy and Phenomenological Research*, *92*(3), 620-644.

Fernández, J. (Forthcoming). The ownership of memories. *The Sense of Mineness*. Eds. García-

Carpintero, M. & Guillot, M. Oxford University Press.

Frise, M. (Forthcoming). Preservationism in the epistemology of memory. *The Philosophical Quarterly*.

Goldie, P. (2012). *The Mess Inside: Narrative, Emotion, and the Mind*. Oxford University Press.

Hamilton, A. (2003). 'Scottish commonsense' about memory. *Australasian Journal of Philosophy*, *81*(2), 229-245.

Holland, R. F. (1954). The empiricist theory of memory. *Mind*, *63*(252), 464-486.

Hopkins, R. (2014). Episodic memory as representing the past to oneself. *Review of Philosophy and Psychology*, *5*(3), 313-331.

Jacoby, L. L., & Whitehouse, K. (1989). An illusion of memory: False recognition influenced by unconscious perception. *Journal of Experimental Psychology: General*, *118*(2), 126-135.

James, S. (Forthcoming). Epistemic and non‐epistemic theories of remembering. *Pacific Philosophical Quarterly*.

Klein, S. B. (2014). Autonoesis and belief in a personal past: An evolutionary theory of episodic memory indices. *Review of Philosophy and Psychology*, *5*(3), 427-447.

Klein, S. B. (2015). What memory is. *Wiley Interdisciplinary Reviews: Cognitive Science*, *6*(1), 1-38.

Klein, S. B., & Nichols, S. (2012). Memory and the sense of personal identity. *Mind*, *121*(483), 677-702.

Kolers, P. A., & Roediger, H. L. (1984). Procedures of mind. *Journal of Verbal Learning and Verbal Behavior*, *23*(4), 425-449.

Lackey, J. (2005). Memory as a generative epistemic source. *Philosophy and Phenomenological Research*, *70*(3), 636-658.

Mahr, J., & Csibra, G. (Forthcoming). Why do we remember? The communicative function of episodic memory. *Behavioral and Brain Sciences*.

Malcolm, N. (1963). *Knowledge and Certainty*. Englewood Cliffs, N.J., Prentice-Hall.

Malcolm, N. (1977). *Memory and Mind*. Cornell University Press.

Martin, M. G. (2001). Out of the past: Episodic recall as retained acquaintance. *Time and Memory: Issues in Philosophy and Psychology*. Eds. McCormack, T. & Hoerl, C. Oxford University Press. Pp. 257-284.

Martin, C. B., & Deutscher, M. (1966). Remembering. *The Philosophical Review*, *75*(2), 161-196.

McCarroll, C. J. (2017). Looking the past in the eye: Distortion in memory and the costs and

benefits of recalling from an observer perspective. *Consciousness and Cognition*, *49*, 322-332.

Michaelian, K. (2011). Generative memory. *Philosophical Psychology*, *24*(3), 323-342.

Michaelian, K. (2011b). Is memory a natural kind? *Memory Studies*, *4*(2), 170-189.

Michaelian, K. (2015). Opening the doors of memory: Is declarative memory a natural kind? *Wiley Interdisciplinary Reviews: Cognitive Science*, *6*(6), 475-482.

Michaelian, K. (2016). Against discontinuism: Mental time travel and our knowledge of past and future events. *Seeing the Future: Theoretical Perspectives on Future-Oriented Mental Time Travel*. Eds. Michaelian, K., Klein, S. B. & Szpunar, K. K. Oxford University Press. Pp. 62–92.

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*. MIT Press.

Michaelian, K., Klein, S. B. & Szpunar, K. K. (2016). The past, the present, and the future of future-oriented mental time travel. *Seeing the Future: Theoretical Perspectives on Future-Oriented Mental Time Travel*. Eds. Michaelian, K., Klein, S. B. & Szpunar, K. K. Oxford University Press. Pp. 1–18.

Michaelian, K. & Sutton, J. (2017). Memory. *Stanford Encyclopedia of Philosophy*. Ed. Zalta, E. N. https://plato.stanford.edu/archives/sum2017/entries/memory/

Perrin, D. (2016). Asymetries in subjective time. *Seeing the Future: Theoretical Perspectives on Future-Oriented Mental Time Travel*. Eds. Michaelian, K., Klein, S. B. & Szpunar, K. K. Oxford University Press. Pp. 38-61.

Perrin, D. and Michaelian, K. (2017). Memory as mental time travel. *The Routledge Handbook of Philosophy of Memory*. Eds. Bernecker, S. & Michaelian, K. Routledge. Pp. 228–239.

Perrin, D., & Rousset, S. (2014). The episodicity of memory. *Review of Philosophy and Psychology*, *5*(3), 291-312.

Robins, S. K. (2016). Misremembering. *Philosophical Psychology*, *29*(3), 432-447.

Robins, S. (2016). Representing the past: Memory traces and the causal theory of memory. *Philosophical Studies*, *173*(11), 2993-3013.

Robins, S. K. (2017). Memory traces. *The Routledge Handbook of Philosophy of Memory*. Eds. Bernecker, S. and Michaelian, K. Routledge. Routledge. Pp. 76–87.

Robins, S. K. (Forthcoming). Confabulation and constructive memory, *Synthese*.

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

Schacter, D. L., Addis, D. R., Hassabis, D., Martin, V. C., Spreng, R. N., & Szpunar, K. K. (2012). The future of memory: remembering, imagining, and the brain. *Neuron*, *76*(4), 677-694.

Schechtman, M. (1994). The truth about memory, *Philosophical Psychology* 7, 3-18.

Shanton, K., & Goldman, A. (2010). Simulation theory. *Wiley Interdisciplinary Reviews: Cognitive Science*, *1*(4), 527-538.

Shepard, R. N. & Chipman, S. (1970). Second-order isomorphism of internal representations: Shapes of states. *Cognitive Psychology, 1*, 1-17.

Squires, R. (1969). Memory unchained. *The Philosophical Review*, 78(2), 178-196.

Suddendorf, T., & Corballis, M. C. (1997). Mental time travel and the evolution of the human mind. *Genetic, Social, and General Psychology Monographs*, *123*(2), 133-167.

Sutton, J. (1998). *Philosophy and Memory Traces: Descartes to Connectionism*. Cambridge University Press.

Sutton, J. (2010). Memory. *The Stanford Encyclopedia of Philosophy*. Ed. Zalta, E. N. https://plato.stanford.edu/archives/spr2010/entries/memory/

Tulving, E. (1985). Memory and consciousness. *Canadian Psychology/Psychologie canadienne*, *26*(1), 1–12.

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

Whittlesea, B. (1997). Production, evaluation, and preservation of experiences: constructive processing in remembering and performance tasks. *The Psychology of Learning and Motivation*, 37, 211-264.

Wittgenstein, L. (1953). Philosophical Investigations. Transl. Anscombe, G. E. M. Blackwell.

Zemach, E. M. (1983). Memory: What it is, and what it cannot possibly be. *Philosophy and Phenomenological Research*, 44(1), 31-44.

Figure 1:

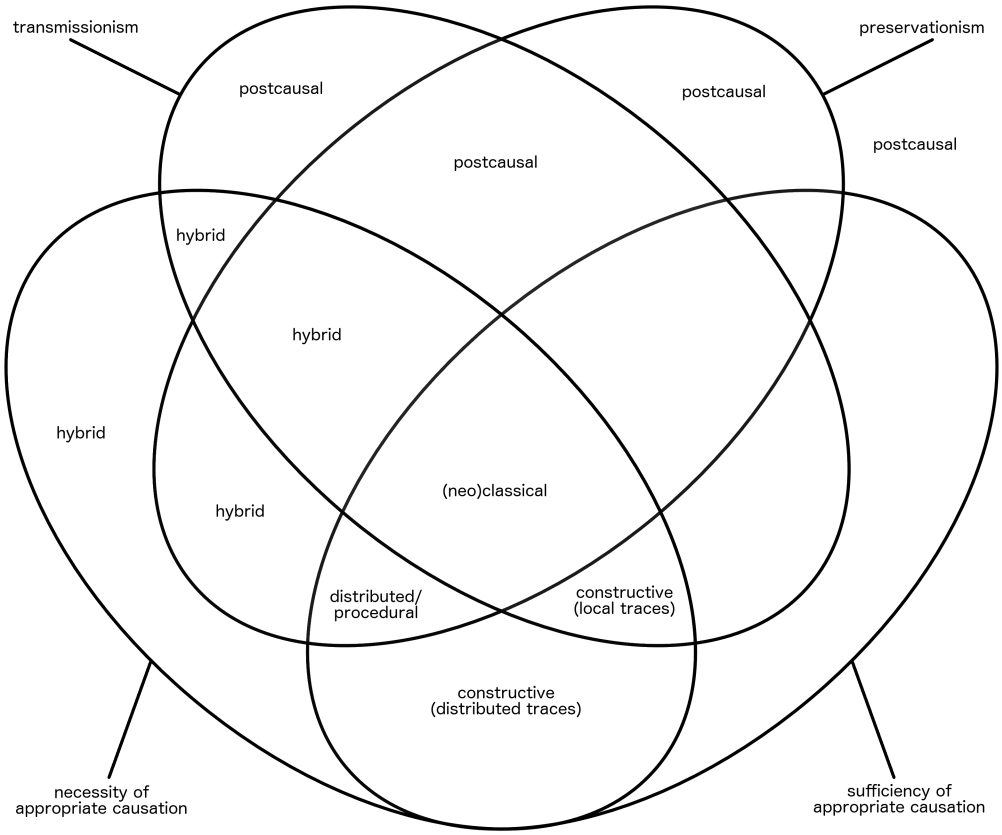
transmissionism
preservationism
postcausal
postcausal
postcausal
postcausal
hybrid
hybrid
hybrid
hybrid
(neo)classical
distributed/
procedural
constructive
(local traces)
constructive
(distributed traces)
necessity of
appropriate causation
sufficiency of
appropriate causation

Figure 1 caption: Relationships among causal and postcausal theories. (*Neo)classical causal theories* (Martin and Deutscher, 1966; Bernecker, 2008, 2010; Cheng & Werning, 2016) maintain that appropriate causation is both necessary and sufficient for remembering and endorse both transmissionism and generationism. *Distributed and procedural causal theories* (Sutton, 1995; Perrin, this volume) agree with (neo)classical causal theories that appropriate causation is both necessary and sufficient for remembering, but their distributed conception of traces leads them to reject transmissionism. *Constructive causal theories* (Michaelian, 2011; Robins, 2016b) likewise agree that appropriate causation is both necessary and sufficient for remembering, but their constructive view of remembering leads them to reject preservationism; the constructive view is compatible with both local and distributed conceptions of traces. *Hybrid causal theories*, including epistemic-causal theories (Debus, 2010) and autonoetic-causal theories (Dokic, 2014; Klein, 2015), depart to some extent from the causal tradition by maintaining that appropriate causation is necessary but not sufficient for remembering; they do not take an explicit stand with respect to transmissionism or preservationism, and the feasibility of the various views in this space remains to be explored. *Postcausal theories*, including the functionalist theory (Fernández, this volume) and the simulation theory (Michaelian, 2016c; cf. De Brigard, 2014a and Shanton & Goldman, 2010), make a decisive break with the causal tradition by maintaining that appropriate causation is neither necessary nor sufficient for remembering. The functionalist theory does not take an explicit stand for or against transmissionism or preservationism. The simulation theory explicitly rejects preservationism but, like the constructive causal theory, might in principle be combined with either a local or a distributed conception of traces and hence might or might not reject transmissionism. *Other theories:* In principle, theories that maintain that appropriate causation is sufficient but not necessary for remembering might be described, but the motivation for such theories is unclear, and none have so far been proposed.