Experimental philosophy of memory

Kourken Michaelian1*, Vilius Dranseika1,2 and Juan Álvarez1

1Centre for Philosophy of Memory, Université Grenoble Alpes, Bâtiment ARSH-1281, Avenue central CS 40700, 38058, Grenoble CEDEX 9, France.
2Interdisciplinary Centre for Ethics and Institute of Philosophy, Jagiellonian University, Kraków, Poland. *Author for correspondence. E-mail: kourken.michaelian@univ-grenoble-alpes.fr

ABSTRACT. Experimental philosophy is now some twenty years old and has a large body of work to its credit. Little of this work focusses directly on memory, but it has, as the philosophy of memory has come into its own over the last several years, become increasingly clear that there are numerous questions about the concept of memory to which the tools developed by experimental philosophers might profitably be applied. By describing a sample of these questions, explaining how and why they might be approached using experimental methods, and providing a snapshot of published and in-progress experimental work, this article makes a case for experimental philosophy of memory.

Keywords: Experimental philosophy; philosophy of memory; memory; philosophical methodology.

Introduction

Experimental philosophy is now some twenty years old and has a large body of work to its credit.1 Little of this work focusses directly on memory, but it has, as the philosophy of memory has come into its own over the last several years, become increasingly clear that there are numerous questions about the concept of memory to which the tools developed by experimental philosophers might profitably be applied. By describing a sample of these questions (section 1), explaining how and why they might be approached using experimental methods (section 2), and providing a snapshot of published and in-progress experimental work (section 3), this article makes a case for experimental philosophy of memory.

1. Philosophy of memory

Though philosophers have long been interested in memory, the philosophy of memory has only emerged as a coherent field of research defined by a recognized set of issues and theories over the last five or so years. We will make no attempt here to provide an exhaustive list of the issues that animate the philosophy of memory,2 but two related debates provide a natural starting point. The first pits causalists (Martin & Deutscher, 1966; Bernecker, 2008, 2010) against simulationists (Michaelian, 2016c, 2021b), with the former holding that genuine memory presupposes an appropriate causal connection—a connection sustained by...
memory trace—to the subject’s experience of the remembered event, while the latter argue that memory is a form of imagination and thus does not require appropriate causation (Debus, 2017; Michaelian & Robins, 2018; De Brigard, 2020; Robins, 2017b). The second pits discontinuists (Debus, 2014; Perrin, 2016; Robins, 2020a) against continuists (Addis, 2020; Michaelian, 2016a), with the former holding that there is a fundamental difference (such as appropriate causation) between remembering the past and imagining the future, while the latter argue that (perhaps because memory is a form of imagination) there is not (Michaelian, Perrin, & Sant’Anna 2020; Munro, 2021; Langland-Hassan forthcoming, Perrin & Michaelian, 2017; Schirmer dos Santos, McCarron, & Sant’Anna forthcoming). Many arguments for the causal theory and for discontinuum appeal explicitly to our intuitions about hypothetical cases of apparent memory—for example, to cases of relearning (Bernecker, 2017a; Michaelian, 2016b; Robins, 2017a, 2020b). Simulationism and continuism, meanwhile, have what appear to be highly counterintuitive implications (McCarroll, 2020). It remains to be determined just how widely-shared the relevant intuitions are, and, as experimental philosophers often focus on eliciting intuitions from a wide range of subjects as a means of investigating the concepts that underlie them, the causalist-simulationist and discontinuist-continuist debates thus represent potentially fertile terrain for experimental philosophy.

In addition to the general questions that figure in the causalist-simulationist and discontinuist-continuist debates, philosophers of memory are interested in a wide range of more specific metaphysical and epistemological questions. One traditional metaphysical question that continues to be debated by philosophers of memory concerns the role of memory in personal identity (Copenhaver, 2017; Heersmink 2017, 2018, 2020ab; Heersmink & McCarron, 2019; Klein, 2013, 2014, 2016; Klein & Nichols, 2012; McCarron, 2019; Nichols, 2017; Roache, 2016; Sattig, 2018ab; Schechterman, 2016, 2021). Another concerns the contents of memory: representationalism, on which a memory represents the event that it is about, has been taken for granted in most philosophy of memory, but recent work has built on research in philosophy of perception to show that relationalism, on which a memory involves a direct relation to the event that it is about, is an alternative worth taking seriously (Andonovski, 2020; Aranyosi, 2020, forthcoming; Barkasi & Rosen, 2020; Moran forthcoming; Sant’Anna, 2020, forthcoming).

One epistemological question that is actively debated by philosophers of memory concerns the factivity of memory (Hazlett, 2010), with intuition-based arguments being offered both for and against the view that memory is necessarily accurate (e.g., Bernecker, 2010). Intuition-based arguments have likewise played an important role in the related debate between alethism, the view that successful memory presupposes truth (accuracy with respect to the remembered event) but not authenticity (accuracy with respect to the subject’s experience of the remembered event), and authenticism, the view that successful memory presupposes both truth and authenticity (Bernecker, 2015, 2017b; De Brigard, 2017; Michaelian & Sant’Anna forthcoming). Another epistemological question concerns the potential immunity to error through misidentification of episodic memory, with recent arguments focussing on the implications of observer memory (see below) for memory’s potential immunity to this form of error (Bermúdez 2012, 2013; Fernández, 2014, 2021b; James, 2021; McCarron, 2018; Michaelian 2021a).

These debates have an a priori, conceptual flavour, and it would be natural for experimental philosophers to begin by intervening here. Other debates within the philosophy of memory are more directly inspired by empirical memory research but nonetheless raise questions about the concept of memory, and there is thus work for experimental philosophers to do here as well. Accounts of kinds of memory based on empirical research on memory systems, for example, appear to challenge traditional philosophical taxonomies (Andonovski, 2018; Cheng & Werning 2016; Colaco forthcoming; Gomez-Lavin, 2021; Klein, 2015; Michaelian, 2011b, 2015; Najenson forthcoming; Werning & Cheng, 2017). Empirically-inspired approaches on which observer memory, in which one adopts a perspective other than that from which one originally experienced the remembered event, can be fully successful similarly appear to challenge traditional

---

1. In addition to explicitly anti-causalist simulation theories (Michaelian 2016c, 2021b), versions of simulationism that may be compatible with causalism have been proposed (De Brigard, 2014; Hopkins, 2018; Shanton & Goldman, 2010). Versions of causalism designed to address problems for the classical causal theory have proliferated in recent years (e.g., Michaelian, 2011a; Perrin, 2021; Sutton & O’Brien forthcoming; Werning, 2020); some of these (e.g., Bernecker 2008, 2010; Debus, 2010) are designed primarily to address conceptual problems and thus may provide suitable targets for experimental philosophy.

2. Fernández’ argument for the functionalist theory of memory (Fernández, 2018, 2019, 2020, 2021a, b), which is distinct from both the causal theory and the simulation theory, appeals explicitly to our intuitions about hypothetical cases and may thus be of interest to experimental philosophers. See Andonovski (2021), James (2021), Robins (2021), and Viera (2021a) for objections to the functionalist theory.

3. Causation is standardly taken to align with discontinuum and simulationism with continuism, but see Langland-Hassan (2021) and Sant’Anna (2021) for more nuanced takes on the relationship between the two pairs of views.

philosophical accounts, which may rule out the possibility of successful observer memory (Lin, 2018, 2020; McCarroll & Sutton, 2017; McCarroll, 2017, 2018; Sutton, 2010). And empirically-inspired accounts of collective (Arango-Muñoz, & Michaelian 2020; Barash, 2017; De Brigard, 2018; Michaelian & Sutton, 2017b, 2019; Seemann, 2019; Sutton, 2018; Theiner, 2013, 2017; Tollefsen, Dale, & Paxton, 2015; Wilson, 2018) and external (Clowes 2013, 2017; Heersmink, 2020ab; Heersmink & Carter, 2020; Heersmink & Sutton, 2020; Michaelian, 2012; Rupert, 2013) memory may likewise challenge traditional accounts, which tend to be individualist and internalist in character.

Before looking at how and why questions such as these might be approached using the tools of experimental philosophy, it is worth noting that there is an emerging focus on memory’s relationships to other cognitive capacities, which has entailed increasing interaction between philosophers who work on memory, many of whom favour a naturalistic approach, and philosophers who work on these other capacities, some of whom continue to rely on traditional a priori methods. In addition to the work on the relationship between memory and imagination cited above in connection with the discontinuism-continuism debate (see also Macpherson & Dorsch, 2018; Berninger & Vendrell Ferrann in preparation), there is emerging work on the relationships between memory and perception (McCarroll, Michaelian, & Arango-Muñoz, 2021) and memory and dreaming (Gregory & Michaelian in preparation). This work can be expected to generate additional questions for experimental philosophy of memory.

2. Why experimental philosophy of memory?

Philosophers of memory may focus on a common set of issues and theories, but there is a deep methodological divide within the field, as there is in the fields, such as philosophy of mind and epistemology, with which it overlaps. Partisans of conceptual analysis, on the one hand, proceed in a largely a priori manner, paying little or no attention to empirical memory research and—more or less explicitly—basing their accounts of remembering on what they take to be widely-shared intuitions. Partisans of naturalism, on the other hand, argue that philosophical accounts of remembering should be compatible with or based on empirical memory research.

There is, of course, no guarantee that accounts based on empirical memory research will align with our intuitions about memory, and a number of partisans of naturalism have argued that they are unlikely to do so. Shanton is particularly explicit on this point:

We have a commonsense picture of how [episodic memory retrieval] works. [...] According to this picture, when we have an experience, we create a detailed file on it. This file is then stored in a mental ‘filing cabinet’ until we want to retrieve it. At that point, the full file is pulled from storage and we read off our memories of the experience from it. Though it’s intuitively appealing, the filing cabinet picture is the wrong picture of [episodic memory retrieval]. (Shanton, 2011, p. 94)

Focussing on the role of memory in choice rather than on episodic memory retrieval, Craver and Rosenbaum remark, in the same vein, that “[…] our armchair intuitions about the role of memory in our lives are a very poor guide to the psychological reality of how episodic memory mechanisms figure in our informed, voluntary choices” (Craver & Rosenbaum, 2018, p. 271). Because they take it to be unlikely that an account based on empirical memory research will align with our intuitions about memory, naturalists have argued that our intuitions should be given little weight in theorizing about memory. Werning and Cheng, for example, maintain that

[...] it would be scientifically and philosophically rather futile to have a notion of episodic memory that, for better or worse, matches our conceptual intuitions, our linguistic practice and perhaps some introspective phenomenology, but does not refer to a natural kind (Werning & Cheng, 2014, p. 968)

while Michaelian suggests that, “[...] while the [causal] theory does a good job of accounting for our intuitions about a broad range of cases, those intuitions themselves may not be a reliable guide to the nature of remembering” (Michaelian, 2016c, p. 80).

While partisans of conceptual analysis generally recognize the force of naturalist arguments, they nevertheless continue to rely on intuition. Bernecker, for example, in the context of his defence of the causalist account of memory, acknowledges that appeals to intuition are problematic but opts to base his...

---

7 This approach is characteristic of most older philosophical work on memory, including Furlong (1951), Von Leyden (1960), Munsat (1966), Smith (1966), Locke (1971), Malcolm (1977), and Warnock (1987) unreferenced, with Sutton (1998) being the first major work to adopt a naturalistic approach.
version of the causal theory on intuition, saying merely that “[…] trying to defend the use of intuitions in the philosophy of memory would ... take us too far afield” (Bernecker, 2010, p. 147). Aranyosi, in the context of his critique of the simulationist account—an account inspired by empirical mental time travel research—of remembering as a form of imagining, adopts a more aggressive posture, saying that

[…] [philosophical intuitions about remembering] are clearly against equating remembering with [mental time travel]. Now, of course, Michaelian might say that his ultimate agenda is to effect a radical change in the philosophy of memory, while leaving the so-called commonsense conception intact, but for that he needs convincing philosophical arguments and thought experiments. What we have so far is an appeal to some empirical findings in neuroscience and cognitive psychology, which won’t convince the philosophers that there really is no deep, important difference between remembering the past and imagining it. (Aranyosi, 2020, p. 579)

Appeals to intuition thus continue to shape much philosophical theorizing about memory.⁸

Partisans of the naturalistic and a priori approaches may disagree about the relevance of the folk concept of memory to philosophical theorizing about memory, but they generally agree on the contours of that concept. Both causalists and simulationists, for example, take it for granted that the causal theory corresponds to the folk concept of memory better than does the simulation theory. There have, however, so far been few empirical studies of the concept of memory. The evidential basis for this consensus is thus slim, and it is here that experimental philosophers stand to make an important contribution to the philosophy of memory.

While experimental philosophy covers a diverse range of projects employing a large variety of empirical methods (Knobe & Nichols, 2017; Sytsma & Buckwalter, 2016; Sytsma & Livengood, 2015), the sort of work that is most relevant in the present contexts consists in investigating our ordinary concepts (for example, the ordinary concept of memory) by collecting study participants’ judgments about scenarios of philosophical interest. Often, this is done by employing the so-called contrastive vignette technique (Reiner, 2019), in which scenarios differ—in the simplest case—only with respect to one tightly-controlled detail, introduced in order to see whether such experimental manipulation will have an effect on study participants’ judgments.

The results that might be produced by applying the methods of experimental philosophy to memory ought to be of interest both to partisans of naturalism and to partisans of conceptual analysis. Those engaged in the project of conceptual analysis should care about the results because they should want to know whether there is a single, widely-shared concept of memory that their theories might in principle track and, if there is such a concept, whether their theories do in fact track it. Those engaged in the naturalistic project of developing a theory of memory on the basis of empirical memory research should likewise care about the results of the experimental approach. Since they take the appropriate object of investigation for the philosophy of memory to be memory itself, rather than the concept of memory, they will not assign the same centrality to these results as do their rivals. But they ought nevertheless to assign them considerable importance: the concept of memory inevitably provides the starting point for empirical memory research and, consequently, for empirically-inspired philosophical theories of memory, and naturalists, too, therefore need to know whether there is a single, widely-shared concept of memory. Without having done the sort of work that experimental philosophers do, we simply cannot be confident, for example, that laypeople and experts share the same concept of memory. Nor can we be confident that, among experts, experts belonging to different professional communities (e.g., philosophy vs. psychology vs. law) share the same concept or that, among laypeople, people belonging to different cultures (e.g., Westerners vs. East Asians) share the same concept. Thus, when Craver (2020), for example, claims that the term ‘memory’ sometimes has an ‘empirical’ sense—a sense that might, in principle, be captured by the simulation theory—and sometimes has an ‘epistemic’ sense—a sense that might, in principle, be captured by the causal theory⁹—we ought, for now, to withhold judgement: before we can reasonably assent to this or other claims regarding the concept of memory, an empirical investigation of that concept is required.

In addition to methods that investigate our ordinary concepts by collecting study participants’ judgments about scenarios of philosophical interest, other methods could and should be used to investigate the concept of memory. For example, whereas experimental philosophy often seeks to determine the contours of our concepts indirectly, by asking subjects to report their intuitions about hypothetical cases, survey methods might be employed in an attempt to determine them more directly, by asking subjects whether they agree or

---

⁸ Andonovski (2021) points out, for example, that Fernández’ (2019) argument for the functionalist theory of memory relies heavily on intuition.

⁹ Note that Craver himself is not very explicit on which theories might best capture the two senses of ‘memory’.
disagree with various claims about memory (Finley, Naaz, & Goh, 2018; Lynn, Evans, Laurence, & Lilienfeld, 2015; Magnussen et al., 2006; Magnussen & Melinder, 2012; Simons & Chabris, 2011). Methods such as topic modelling likewise hold considerable promise (see Malaterre, Chartier, & Pulizzotto, 2019; Malaterre, Pulizzotto, & Lareau, 2020; Malaterre, Lareau, Pulizzotto, & St-Onge, 2021; Weatherson, no date). Our focus here, however, is squarely on methods aimed at collecting judgments about cases. The following section demonstrates the need for more such work by reviewing published and in-progress experimental work on the concept of memory, showing that, while it begins to shed light on some of the questions outlined in section 1, there is much more to be done.

3. Existing work

There are a number of experimental studies that address memory not in its own right but rather in the context of other philosophical problems, such as personal identity. In this context, the question at issue is most often whether the folk apply the memory criterion of personal identity—that is, whether they think that continuity of memory is what allows a person to persist though time—or whether they instead employ some other criterion, such as bodily continuity or continuity of moral character (Nichols & Bruno, 2010; Strohminger & Nichols, 2014; 2015; Tierney, Howard, Kumar, Kvaran, & Nichols, 2014; Berniūnas & Dranseika, 2016; Woike, Collard, & Hood, 2020). The experimental philosophy literature on personal identity tends to provide study participants with stories of people undergoing various types of transformation (Dranseika, Dagys, & Berniūnas, 2020) in an attempt to ascertain the criteria on which people rely in deciding whether the post-transformation individual is ‘the same person as’ the pre-transformation individual. Interestingly, while continuity of memory does indeed seem to matter for folk judgements of personal identity over time, continuity of moral character appears to matter more. This work should be read in conjunction with literature in the cognitive science of religion investigating the role of autobiographical beliefs about past lives in explanations of reincarnation beliefs (White, 2015; 2016a; 2016b; White, Kelly, & Nichols, 2016; see also Dranseika forthcoming). The ethnographic record suggests that cultures that have both reincarnation beliefs and practices of reidentification of reincarnates tend to base the latter practices on memory tests or bodily marks. This literature tries to understand the features in virtue of which memory tests (e.g., whether the child will be able to recognize items that belonged to the deceased Buddhist monastic) are taken to provide convincing evidence that an individual has lived past lives. ¹⁰

In addition to studies that address memory in the context of personal identity, there are studies—many of them in progress or not yet published—that address memory in its own right. Álvarez et al. (unpublished data) ask whether the folk concept of memory is compatible with relearning, with preliminary results suggesting that cases of relearning are regarded neither as clear instances of genuine memory nor as clear instances of merely apparent memory. Given the centrality of appeals to intuition about cases of relearning in arguments for the causal theory of memory, these results may be problematic from a causalist point of view.

Furthering the work on personal identity described above, Dranseika (forthcoming) looks at whether people treat memory claims as evidence of personal identity, focussing, in particular, on memory claims regarding past lives. When potential memories of past lives are considered, the characteristic of verifiable memory claims that calls for an explanation—and that is sometimes explained in terms of reincarnation—is the possession of otherwise hard-to-obtain knowledge about past events rather than whether the memory claim is presented as based on personally remembering the event.

Dranseika (2020) looks at both factivity and the relationship between memory and dreaming. His data suggest that ordinary attributions of ‘to remember’ and ‘to have a memory’ can violate the factivity constraint (in, for example, memories of events that in fact did not happen—either memories of dream events or completely artificial implanted memories). Interestingly, his data also suggest that memories can violate the previous awareness condition (in, for example, science-fiction style memory transfer scenarios), a condition that causalists accept but that simulationists reject. Dranseika, Robins, and Michaelian (unpublished data) look more closely at factivity, attempting to determine whether successful memory presupposes truth, authenticity, or both. Their data suggest that authenticity matters more than truth in ascriptions of

¹⁰ Memory has also been discussed in experimental studies of knowledge, though the links with philosophy of memory are less pronounced here than they are in experimental studies of personal identity. Turri (2016), for example, uses memory as one possible way of manipulating the reliability of a source in studying attributions of knowledge. Powell, Horne, and Pinillos (2014; Powell, Horne, Pinillos, & Holyoak, 2015) introduce an experimental paradigm, semantic integration, that employs memory tasks to investigate various concepts, including the concept of knowledge. And Kosuch and Nichols (2011) study folk intuitions about the transparency of introspection by using vignettes featuring descriptions of classical experiments involving memory tasks.
remembering. For instance, study participants are more inclined to treat false but authentic images as memories than true but inauthentic ones. Surprisingly, then, it may be that neither alethism nor authenicism is right with respect to the folk concept of memory.

Dranseika, McCarril, and Michaelian (2021) provide evidence that the folk concept of memory does not preclude observer memory. Study participants were willing to agree that visualizations of past events from an observer perspective are instances of remembering. Their study also found that observer memory tends to be regarded as slightly less accurate and more distorted than field memory, in which one adopts the perspective from which one originally experienced the remembered event. Overall, the data suggest that philosophers may have overestimated the conceptual importance of the difference between field and observer memory. Finally, Dranseika, McCarril, Michaelian, and Sakuragi (unpublished data) examine potential differences between English- and Japanese-speakers with respect to the collective dimension of memory, asking whether Japanese-speakers assign more weight to agreement of a memory with the memories of other group members and English-speakers more weight to the accuracy of the memory.

Many of the studies reviewed in this section are in progress or not yet published, and more work will be required to confirm the results of those that are already published. This brief review should, however, suffice to demonstrate that experimental research can indeed shed light on questions that are of interest to both naturalistic and a priori philosophers of memory.

**Conclusion**

Our goal in this article has been to make a case for experimental philosophy of memory. We have described a sample of the questions about the concept of memory on which the methods of experimental philosophy might be trained, explained how and why those questions might be approached using those methods, and reviewed the work that has begun to do so. We thus hope to have made clear that there is both room for and a need for a great deal of additional experimental research on the concept of memory. While there is some work on the necessity of appropriate causation (specifically, on the compatibility of relearning with remembering), the role of memory in personal identity, factivity, authenticity vs. truth, observer memory, collective memory, and the relationship between memory and dreaming, more work ought to be done on each of these questions, and no work at all has been done on representationalism vs. relationalism, immunity to error through misidentification, kinds of memory, external memory, or the relationships between memory and imagination and between memory and perception.

These questions are those with which we began and are, again, by no means meant to provide an exhaustive list of the questions that are currently debated by philosophers of memory. A qualitative review of recent publications reveals a long list of other questions that might be in principle be approached experimentally, such as the nature of memory traces, the phenomenology of memory, metacognition in memory, forgetting, the role of emotion in memory, the relationship of memory to the self, the nature of procedural memory, memory and narrativity, memory and time, reference in remembering, and the relationship between remembering and knowing. To provide just a bit more detail on the last of these examples, Bernecker (2007) argues that remembering does not entail knowing, whereas Moon (2013) maintains that it does; both authors appeal explicitly to intuition, as do others who have defended the same views. Data collected by David Rose (unpublished manuscript) suggest that ascriptions of remembering entail ascriptions of knowing and hence that Moon may be right, but more work will be required before this can be established with confidence.

There are, in short, numerous questions about the concept of memory to which the tools of experimental philosophy might profitably be applied, and we therefore close this article by calling for experimental philosophers to join forces with philosophers of memory to shed additional light on the concept.

**Acknowledgements:**

Thanks to André Sant’Anna for the invitation to write this article. Kourken Michaelian was supported by the French National Research Agency in the framework of the “Investissements d’avenir” program (ANR-15-
IDEX-02). Vilius Dranseika was supported by the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation program (grant agreement 805498).

References


Colaço, D. (Forthcoming). What counts as a memory? Definitions, hypotheses, and “kinding in progress.” *Philosophy of Science.*


Schirmer dos Santos, C., McCarroll, C. J., & Sant’Anna, A. (Forthcoming). The relation between memory and imagination: A debate about the “right concepts”. In A. Sant’Anna, C. McCarroll, & K. Michaelian (Eds.), *Current Controversies in Philosophy of Memory*. New York, NY: Routledge.


