Penultimate draft, forthcoming in Andonovski, McCarroll, Michaelian, & Sant'Anna (Eds.) *Philosophy of Mind and Cognition*. Palgrave.

Memory

1 Introduction

In his 'semiautobiography', Luis Buñuel writes: "Life without memory is no life at all. [...] Our memory is our coherence, our reason, our feeling, even our action. Without it, we are nothing" (1982: 4–5). We get closest to grasping what life without memory would be like when we consider the densely amnesic. Clive Wearing was an accomplished musicologist and musician before a viral infection led to encephalitis, causing particular damage to regions involved in memory. Upon regaining consciousness, he recalled enough to engage in short conversations about parts of his life and some areas of general knowledge. He was able to sing and play music, and remained in love with his wife. Yet he was unable to retain new information for more than half a minute. He would repeatedly conclude he had newly regained consciousness, announcing this in conversation or sometimes jotting it down in a diary, crossing out previous declarations none of which he could recall having made:

> 8:07am: I AM ~~totally, perfectly~~ awake. ~~1st Time~~.
> 8:31am: Now I am ~~really, completely~~ awake. (~~1st Time~~)
> 9:06am: Now I am ~~perfectly, overwhelmingly~~ awake. (~~1st Time~~)
> 9:34am: Now I am ~~superlatively, actually~~ awake. (~~1st Time~~) (Wearing 2005).

Wearing cannot read books or follow extended conversations. He has frequently described his plight as 'like being dead'.

Few aspects of the mind cut as deep as memory, and its pertinence for questions about self and identity, knowledge, and moral responsibility indicates its philosophical centrality. For instance, to what extent is Wearing responsible for actions he cannot recall, and what implications does his lack of diachronic agency have for selfhood? Yet it is only in the last decade or so that memory has become a sustained focus for analytic philosophy in the way perception or language have long been. Spurred on by recent developments in the sciences, memory is now a vibrant area of philosophical inquiry.

This introductory chapter provides an overview of key issues in the philosophy of memory. §2 starts at the most general level with the question of what memory is and lays

some conceptual groundwork. §3 surveys traditional approaches to memory, highlighting a perennial focus on a form of remembering partly constituted by imagery. §4 turns to three major empirical developments and their philosophical implications: the constructive character of memory, the multiple memory systems framework, and episodic memory as a form of mental time travel. §5 briefly illustrates two topics of current interest: mnemic reference and memory errors. These topics are live and can be expected to illustrate the fruitful interplay between philosophy and the sciences that characterises the field.

2 What is memory?

It is possible to conceive of memory broadly, as any means by which an individual's behaviour can be influenced and shaped by their past. Along these lines: "memory can be defined as experience-dependent modification of internal structure [...] that alters the way the system will respond to stimuli in the future" (Baluška & Levin 2016: 2). However, this definition would include many things we might not intuitively recognise as memory. For instance, while a rat's ability to learn the optimal route through a maze after several trials would constitute an exercise of memory on this definition, it seems that retinal damage from staring at the sun would, too.

It is also possible to conceive of memory narrowly. As Klein puts it, "'[m]emory' is a term that is (or should be) reserved for those experiences directly felt to be a reliving of the circumstances from which they were acquired" (Klein 2015: 6). But this definition would *exclude* many things we might intuitively wish to recognise as memory. For example, while your ability to remember your first day at school would constitute an exercise of memory on this definition, it is not clear that your ability to recall that Yaoundé is the capital of Cameroon, or to whistle your favourite Beatles song, would also be included.

Each proposed definition has its own distinctive, scientific aims. It would be unwise to defer to our intuitions on how to conceptually carve up the mind once and for all. Which among the various clusters of phenomena we ought to call 'memory' is a live and important metatheoretical question (see Colaço (2022) & Feest (2025) for relevant discussion). What is certainly true is that the phenomena of memory, studied across disciplines and historical traditions, are complex and heterogeneous. This chapter will focus on memory phenomena most closely related to the concerns of the Western philosophical tradition.

Alongside decisions about how broad to conceive of memory, the term 'memory' also exhibits a degree of ontological polysemy. That is, it can be used to refer to a *capacity* to encode, store, and retrieve information; a hypothetical *store* of information; the body of information *in* that store; a particular *item* of information in that store; an instance of *recalling* a fact or *remembering* a particular event, object, or place; and more besides. When we say someone has a good memory, we typically mean the first thing on this list (a *capacity)*. When we say someone has a vivid memory of their first day at school, we typically mean the fourth (a *state* of memory) or fifth thing (an *occurrence* of remembering). Though I will disambiguate where appropriate, we will be interested primarily in occurrences of remembering and the underlying memory states of which they involve the activation (albeit perhaps in an attenuated sense (§4.1)).

3 Traditional philosophical approaches

According to Aristotle's *De Memoria et Reminiscentia* (ca. 350 BC), memory states, which follow experience or learning, are a kind of imprint upon a bodily organ. Subjects able to excite this imprint and entertain the corresponding image (*eikôn)* undergo occurrences of memory. These picture-like images of remembering are said to be not only a likeness of the remembered particular but are causally derived from the subject's past awareness of it. For example, in remembering one's first day at school, one entertains an image—perhaps of the classroom, or of one's mother waving goodbye at the gates—that is a likeness of what was experienced and owes its existence to that experience. Moreover, in remembering, the subject understands that they previously encountered or learned the remembered thing. Objects of memory are always *past*, for Aristotle, and range from facts, events, and individuals to things like lists of items and intervals of time.

Aristotle's account already exhibits sophistication, anticipating notions we shall encounter later: *memory trace*, *memory causation*, *encoding*, *storage*, *retrieval*, and a *previous awareness condition*. Inevitably, there are idiosyncrasies and oversights. As Aristotle conceives of memory, it is the activity of a *perceptual* faculty, and "[m]emory, even the memory of objects of thought, is not without an image" (449b30; in Sorabji (1972: 49)). But we might again wonder whether we ought to make room for a subject's memory that Yaoundé is the capital of Cameroon without their entertaining imagery at all.[1] Moreover,

[1] Importantly, Aristotle's notion of *reminiscence* or intellectual memory does not play this sort of role.

we might note that in this latter sort of case the object of memory (or its truth) need not be confined to the past.

While Book X of Augustine's (1795) *Confessions* (ca. 400 AD) characterises memory as a 'storehouse for countless *images* conveyed by the senses', he distinguishes elements of the storehouse, and occurrences of memory, corresponding to facts or laws of mathematics themselves ('none of which were imprinted by any bodily sense' (§12)). These are independent of their expression in natural language or via sensory mental imagery. This improved view avoids Aristotle's implication that memory requires mental imagery.

3.1 The search for kinds of memory

In Augustine we can start to make out a tendency that strongly characterises the study of memory: the aim to identify *kinds* of memory. In pursuing this, philosophers have a habit of proposing taxonomies using terminology that does not clearly overlap with precedents. Across the 20th Century: Russell (1921/2005) distinguishes true memory and habit memory; Broad (1925) distinguishes perceptual and non-perceptual memory (the latter including propositional memory but not 'motor capacities'); Furlong (1951) distinguishes retrospective and non-retrospective memory (the latter including remembering-that and remembering-how); Ayer (1956) distinguishes recollective, propositional, and habit memory (suggesting all may be reducible to the latter); Malcolm (1963) distinguishes personal, perceptual, and factual memory (non-exhaustively); and, Locke (1971) distinguishes personal, factual, and practical memory. This is only a sample. Philosophical study of belief, desire, and perhaps even perception, has not been witness to such a dizzying array of distinctions. It is easy to sympathise with Brewer (1996), who concludes that "the total lack of convergence [...] shows that all these attempts were failures" (21).

It is important to note that these taxonomic differences intersect with differences in project. For instance, the 'linguistic/conceptual turn' in analytic philosophy saw many become largely interested in our ordinary terms/concepts (e.g., Zemach (1968)). Proponents of this *analytical approach* sought to disambiguate and define the ordinary concept of memory (or term 'memory'). For instance, Martin & Deutscher (1966) claim to show that there is a "logical connection between [the sentences] 'A's past observation of *X* is causally related to his present representation of it' and 'A remembers *X*'" (175). Similarly, Anscombe (1981) suggests it is *analytic* (i.e., part of the meaning of 'memory') that "[t]he

original witnessing of a remembered event is a cause of any present memory (even a false memory) of that event" (130).

Some philosophers expressly interested in memory itself—as opposed to an ordinary concept or term—have nonetheless proposed taxonomies grounded in linguistic distinctions. Proponents of this *semantic ascent approach* might note that the verb 'remember' is highly flexible and can take many different arguments: 'that 311 is prime'; 'her 18th birthday party'; 'her father's rage'; 'his first bicycle'. Correspondingly, Bernecker (2008) distinguishes fact, event, property, and object memory as kinds of memory voiced by such constructions. Yet distinctions grounded in features of natural language memory reports, though useful for some purposes, seem unlikely to limn the mind's real structure.[2]

Most philosophers interested in memory itself distinguished kinds of memory largely via introspection. With few exceptions, such *introspective approaches* echo Aristotle and Augustine in recognising a special form of memory that crucially involves *sensory mental imagery*. For Russell (1921/2005: 155), true memory "demands [...] an image". For Furlong (1951), retrospective memory involves imagistic reproduction of a past state of mind. For Malcolm (1963), "perceptual memory [essentially] involves mental imagery" (207). And, for Locke (1971), personal memory "essentially involves mental imagery" (80). Such imagery is characterised—visuocentrically—as 'dim', 'sketchy', 'schematic', or 'washed out'.

3.2 An experiential kind of remembering?

Some remain centrally interested in this putative explanandum revealed by introspection. Martin (2015) suggests that "in this kind of case, the imagistic component seems to constitute the remembering" (34), and Hopkins (2024) that "the imagistic state of mind *is* the memory" (152). Some with this explanatory target in mind have characterised it as the acquisition, retention, and activation of a distinctive sort of knowledge concerning the phenomenal character of past experiences (e.g., Hoerl (2018), Soteriou (2013)).

[2] 'Semantic ascent' strategies continue to be explored (Liefke & Werning 2024). Those suspicious of the methodology may still hope for approximations of linguistic markers for kinds of memory in order to assist execution of paradigms such as the Autobiographical Interview (Levine et al., 2002).

An historical preoccupation among those interested in this form of remembering has been 'memory imagery' and its relation to the object of conscious awareness. Typically, such a philosopher notes that the image cannot be *what* one remembers, for the image is not past. This is widely followed by rejection of the *indirect realist* view on which, though the memory is of the past event, the object of one's awareness is a present image (cf. Locke (1690)). This view was fiercely rejected by Reid (1765): "Philosophers indeed tell me, that the immediate object of my memory [...] is not the past sensation, but an idea of it, an image"; "I beg leave to think with the vulgar, that when I remember the smell of the tuberose, that very sensation which I had yesterday [...] is the immediate object of my memory" (32–33).

Direct realists agree with Reid that remembering can afford awareness of past particulars themselves. Yet, like philosophy of perception, a live debate concerns the nature of this awareness. For *relationists* (Debus 2008), such awareness consists in the subject bearing a conscious relation to the thing remembered, a relation of a sort that can only obtain in virtue of the (atemporal) existence of its relata. As a result, successfully remembering an event affords *experience* of a fundamentally distinct kind to merely seeming to remember an event that never occurred. According to *representationalists* (Martin 2019), in contrast, recollective awareness consists in the subject bearing a conscious representational relation to the thing remembered, thus one which is not grounded in the existence of its relata.

A neglected question in this debate is whether motivations analogous to those that drive relationism about *perceptual* awareness transfer over to the case of recollective awareness. Barkasi & Sant'Anna (2022) suggest relationism can explain phenomenological features that representationalism cannot so well account for: "[a]s you recall the party, this experience [...] may be less 'vivid' than the original [...], but it still introspectively strikes you *as if you're perceiving the event again"* (12). Though it is common to describe the phenomenology of remembering in terms of 'reliving', this may be too strong. Indeed, Martin (2015) highlights concern that relationism about recollective awareness (in conjunction with the perceptual) may leave us with insufficient resources to explain manifest phenomenological differences between perceiving and remembering.

A further issue, cutting across the representationalist-relationist debate, concerns what the strict object of awareness is in this form of remembering. In the passage quoted

earlier, Reid suggests it is the 'very sensation' he had. Yet elsewhere in the same passage he suggests it is 'the fragrance smelled'. This suggests an ambiguity about whether the object of awareness is a past *experience* or a past entity *experienced*.[3] On one version of the former view, recollective awareness consists in imagistically representing one's previous (relational) awareness of the remembered event (Martin (2019); Hoerl (2018)). I have argued elsewhere that *experientialism*—the thesis that to remember *x* is to remember *experiencing x*—is difficult to reconcile with cases in which we remember an object or place without remembering any event in which we experienced it (Openshaw 2022).

Admittedly, experientialism provides a tidy explanation for a key feature of this apparent form of remembering, one ubiquitously observed and expressed best by Reid.

> Things remembered must be things formerly perceived or known. I remember the transit of Venus over the sun in the year 1769. I must therefore have perceived it at the time it happened [...]. Our first acquaintance with any object of thought cannot be by remembrance (1785: 305).

This has come to be known as the 'previous awareness condition'. An explanation of why it seems manifestly true (Hoerl (2018: 320); Martin (2019: 117)) may be that it is entailed by experientialism plus the truism that one can only remember what has actually occurred.

Yet whether the previous awareness condition, experientialism, or anything else, is *constitutive* of some form of remembering depends on how we demarcate the form in question. This section has explored one perennial explanandum among philosophers of memory: a form of remembering partly constituted by conscious imagery of a certain sort. Although some maintain special interest in this explanandum, we will see that it is very difficult to map this onto the concepts and concerns of the recent, naturalistic landscape.

4 Major empirical developments and their philosophical implications

In the last decade or so, philosophy of memory has been driven by a deep and prevailing methodological shift. *Methodological naturalists* investigate mental aspects of the world in much the way one would investigate others, attempting to construct explanatory theories

[3] See also Aristotle (Sorabji 1972: 7–8, n.).

with a view toward integration with the natural sciences.[4] There may be no guarantee that the theoretical understanding reached will harmonise with common-sense or folk psychology, much as physics is not beholden to folk physics. In sharp distinction to traditional approaches (§3.1), linguistic and conceptual concerns have no special place, and introspection and linguistic usage have no special role.

For naturalists, definitions of memory reached by armchair introspection and thought experiment are, at best, a starting point. As Michaelian has put it, any appeal to intuition is "provisional at best", for while intuitions may "provide insight into our ordinary concept of remembering, that concept itself fails to correspond to the nature of [...] remembering" (2016: 61). Indeed, to what extent they even constitute a starting point is a matter of temperament. Some will be amenable to redefining their questions and aims in light of scientific developments. And this may lead to considerable shifts in the very targets of theorising.[5] Others may think it desirable, however far science progresses, to have a means of translating developments into an idiom via which they can be grasped as answering the questions we set out to answer (e.g., 'What is *remembering?')*, perhaps reconciling what Sellars (1997) called the scientific and manifest images.

Whatever the details, perhaps the unifying attitude among contemporary theorists is that we ought not sharply distinguish inquiry into the *nature* of memory ("what memory is in itself [...], however it may operate" (Locke 1971: 1)) from inquiry into the *workings* of memory. In rigidly imposing such a distinction, one will tend to treat armchair inquiry as a privileged means of categorising and understanding mental phenomena. As we have seen, this is likely to get us only so far. If, instead, empirical inquiry can be seen as directly related to the project of explaining what memory is, some form of methodological naturalism will follow. Mature developments in the science of memory will have immediate implications for philosophical theorising.

This section presents three key waves in the science of memory that have most strongly shaped contemporary philosophical theorising.

---

[4] To be distinguished from *metaphysical* naturalism, a substantive thesis about what there is: namely, only those entities posited by physics and/or other appropriately 'basic' sciences. For a thorough discussion of naturalism in philosophy of memory, see Andonovski & Michaelian (2024).

[5] For a good example, see Hoerl's (2022) accusation of eliminativism.

4.1 Memory as constructive

Ebbinghaus is widely recognised as having taken psychological study of memory beyond introspection to experimentation. Using himself as subject and nonsense syllables (e.g., *leq, muv)* as stimuli, Ebbinghaus (1885) demonstrated systematic relationships between conditions of learning and amount learned, as well as the exponential rate at which information retention tends, if not rehearsed, to decline over time (his 'forgetting curve').

Ebbinghaus continued to trade in the ancient metaphor of stored mental 'inscriptions' that are later 'excited'. At a fair degree of abstraction, this has since come to be known as the E-S-R model, decomposing the temporally extended process of memory into three basic, functionally distinct stages: encoding, storage, and retrieval. The model itself is neutral on how encoding occurs, where information is stored, etc. Yet, soon after Ebbinghaus, developments in the early psychology of memory led to key revisions in its character.

Bartlett (1932) used complex, cultural materials in experiments on memory. His suggestion was that memory is bound up with an 'effort after meaning', errors often being functions of subjects' background knowledge. Bartlett gave subjects a folk tale from an unfamiliar culture, with supernatural elements and an unusual structure. Re-tellings by participants saw the story change in ways that conformed more to their own cultural expectations (e.g., 'hunting seals' became 'fishing'), a process Bartlett called *rationalization*. He posited internal representations—schemas—that encode abstracted commonalities of thematically related experiences and influence subsequent memory. It is now thought that schemas (e.g., scripts, frames, stereotypes) enhance memory, both because schema-inconsistent information is better *encoded* and because schema-consistent 'guesses' by memory at *retrieval* are, in everyday life, frequently correct.

These facts are enough to drive a wedge between two general pictures of memory and of the E-S-R model. On the *archivalist* picture, the basic task of memory is to fixedly encode a given content and store it to afford its future retrieval as-was. In other words, remembering is a passive process in which a representation with pre-existing content is retrieved. Though some content may be lost, the overarching aim is to preserve while preventing foreign intrusions. On the *constructivist* picture, in contrast, the task of memory is more subtle, perhaps being to equip the organism with relevant, accessible information

at an efficient cost and for myriad uses. Encoding itself is selective, storage dynamic, and retrieval combinatorial.

Subsequent influential studies—e.g., Loftus et al.'s (1978) *misinformation effect* or the so-called *DRM effect* (Roediger & McDermott 1995)—have been taken to suggest that "any contrast between reproductive and reconstructive memory is ill-founded; all remembering is constructive in nature" (Ibid: 812). What this means is crucially *not* that memory is unreliable. Visual illusions result when adaptive processes, which serve humans well in their usual ecological niche, go astray in artificial contexts. They do not show that one cannot trust one's senses. In the same way, these constructive processes are born of expedience and efficiency. The DRM effect may result from subjects' implicit self-generation of the lure during the study phase, later misremembered as something they heard. But, in everyday life, the effects of such processes will often be accurate representations that guide our actions toward success.

The emerging picture of memory as constructive has shaped epistemological study of memory. Insofar as "achieving the central epistemic goal of memory involves making new content" (Aronowitz 2018), it is natural to conclude that one can remember that *p* even if one never previously entertained that *p* (on *content generationism* see, e.g., Michaelian (2011)), and that one can be *justified* in believing that *p* via memory in cases where this is not merely a matter of having preserved justification from some other source (on *epistemic generationism* see, e.g., Tooming & Miyazono (2024)).

4.2 Memory systems as kinds of memory

Philosophers were relatively quick to conclude that memory is not *one* kind of thing. But, as we saw in §3, the contours of its decomposition were long disputed. In answer to 'What is memory?', naturalistic approaches find their basis in the *systems* of cognitive psychology. Memory systems, on this view, are instantiations of the real kinds of memory.

Each system of memory is thought to employ representations of a certain sort, transforming these according to distinctive operational rules, localised in some more or less specialised brain region. Albeit distinct, systems causally interact and frequently overlap in their constituent parts and processes. It is, as a general rule, through cooperative interaction that these systems produce the "introspectively apprehensible and objectively

identifiable consequences of learning and memory" (Schacter & Tulving 1994: 3). For this reason, conclusions about systems—as opposed to their effects—will typically require a convergent range of evidence: neuropsychological dissociations, functional dissociations, neuroimaging evidence, etc.

Orthodoxy divides the systems of human memory as follows (Squire 2004). Short-term memory involves the retention of small amounts of information for up to around 30 seconds. *Long-term* memory involves the retention of information of myriad kinds and combinations. It has practically unlimited capacity and what is stored may persist for a lifetime. Long-term memory can be helpfully divided into the declarative (or explicit) and non-declarative (or implicit). The former is accessible and articulable by the subject (e.g., recalling the capital city of Cameroon). The latter is not, and there is no corresponding gap between retrieval and expression in behaviour (e.g., motor skills).

Declarative memory is traditionally subdivided into *episodic* and *semantic* memory. We can gloss their distinction by saying episodic memory affords qualitatively rich representation of spatiotemporal circumstances (e.g., remembering one's 18th birthday party), whereas semantic memory affords knowledge divorced from the circumstances of its acquisition (e.g., recalling that Charlemagne was Holy Roman Emperor, or visualising the map of Italy).[6] Following Endel Tulving, episodic memory is conceived of as a system specialised in encoding, storing, and retrieving traces formed upon experience of specific events. It is also, Tulving (1985) supposed, necessary for a special form of 'autonoetic' consciousness of one's past and identity across time.

A striking case that seemed to reveal the episodic-semantic division is that of patient K.C. Following a motorcycle accident that caused severe medial temporal lobe damage, especially to hippocampal areas, K.C. suffered dense episodic and semantic anterograde amnesia. However, his retrograde amnesia seemed selectively episodic. According to Tulving, K.C.'s knowledge of his own past had "the same impersonal experiential quality as his knowledge of the rest of the world" (1985: 4).

[6] Despite the term 'semantic', it is not (any longer) thought to be tied to information concerning, or expressible in, language. It also cannot be construed as a body of amodal, propositional information.

The naturalist's working hypothesis is that episodic memory comprises a *natural kind* of memory. This means that the construct *episodic memory* denotes a class of processes that share many properties that subserve inductive practices (and which is maximal with respect to processes that share these properties because of a common mechanism). Crucially, the same cannot be said for personal memory, recollective memory, retrospective memory, etc. (§3.1). The point is that classifying entities by their introspective or superficial character will often lead to bad inductive inferences. Biological generalisations over the class of weeds, for example, will be weak because its members do not share a relevant underlying causal structure. In the same way, naturalistic theories of memory will be better to the extent that they classify in ways that optimise inductive potential. And many see memory *systems* as the basic units by which to secure this potential.[7]

A cautionary note before continuing. Philosophers sometimes talk as if 'episodic memory' is psychology's term for the principal phenomenon in which philosophers have long been interested (§3). It is often supposed that everyday experiences of remembering birthday parties etc. comprise a *kind of remembering*—'episodic remembering'—a unitary phenomenon related to a single system (e.g., Michaelian & Sutton (2017)). This assumes both that there *is* a unitary phenomenon on which philosophers were converging, despite their terminological differences (§3.1), and that psychological inquiry concerns that same thing. But if "[o]ne of the most fundamental assumptions of neuropsychology is that the carrying out of any cognitive task [...] is codetermined by the operations of a number of different systems" (Tulving 1991: 22–23), care must be taken. Moving seamlessly between empirical claims about an episodic system and claims about the putative subject matter of traditional debates is inferentially risky: there may be instances of the phenomenon that are independent of episodic memory, and there may be episodic activities that do not support instances of the phenomenon.

4.3 Memory and mental time travel

While Tulving had conceived of episodic memory as a "past-oriented memory system" that "allow[s] one to re-experience [...] one's own previous experiences" (2002: 5), the landscape soon shifted. It had struck Tulving (1985) as interesting that patient K.C.'s episodic amnesia coincided with an apparent inability to imagine events in his personal future. In

[7] See Ch. 2 of Michaelian (2016) and Cheng & Werning (2016) for more on memory and natural kinds.

2007, findings emerged from several labs seeming to confirm that the neural regions active in remembering specific events in one's past substantially overlap with those active in imagining possible events in one's future (*inter alia,* Hassabis et al. (2007)). These findings fuelled speculation that what had been previously conceived as episodic memory was in fact one aspect of a larger system, or 'core network', for constructing personal event-representations. Schacter and Addis's *constructive episodic simulation hypothesis* proposed to explain both the above neural overlap and the strikingly constructive character of remembering one's past: the system is not *for* remembering, but for 'simulating episodes that might occur in our future' (2007: 778). As the function of an episodic system, mental time travel 'from the present to the past' is not special.

Thus, according to a dramatic and influential way to interpret post-2007 mental time travel research, *there is no* episodic memory system. Rather, there is a system with a more general function: producing rich simulations of past, future, or possible events. As one might say, episodic memory does not denote a *maximal* class underpinned by a uniform causal mechanism, and so it is not a natural kind. As Michaelian has put it,

> The episodic memory system is in reality a general episodic construction system, designed to draw on information originating in past experience to simulate possible episodes. [...] remembering is not different in kind from other episodic constructive processes (2016: 103).

These developments have rekindled an old debate, concerning the relation between remembering and *imagining*. Hobbes declared that "Imagination and Memory, are but one thing" (1651/2011: 16). While Hume (1739/1978) attempted to distinguish the two, the difficulty he had in doing so by considering only their introspectible characteristics suggests key similarities. In one way or another, comparisons between remembering and imagining have been ubiquitous in philosophy. In psychology, too, Bartlett's conclusion that memories are not 'fixed and lifeless' was enough to prompt the question (1932: 312), "if we say that memory is itself constructive, how are we to differentiate it from constructive imagination"? A rich array of debates has emerged in the last decade, aiming to address *inter alia* whether remembering is, in the same way as imagining, an action (Hopkins 2018), skilful (Goldwasser 2023), or controlled (Sant'Anna 2023), and whether there are differences in attitude (Langland-Hassan 2022) or direction of fit (McCarroll 2022).

While disputes about the distinction between remembering and imagining are far from new, the core debate—*discontinuism* versus *continuism* (Perrin 2016), as it has come to be known—stems from, and will ultimately be decided by, questions about memory systems and their component processes. Continuists, such as De Brigard (2014), argue that the episodic system (and any of its component processes) is not for remembering *per se* but for, e.g., rich hypothetical thought about relevant possibilities. To that extent, "there is no deep, categorical difference between 'remembered' and 'imagined' event representations" (Mahr 2022: 227). Remembering may involve retrieval of traces, but the same retrieval processes are active in imagining (Langland-Hassan 2022b). Discontinuists, such as Werning (2020), argue that the episodic system *is* a *memory* system, with some specialised processes, such as retrieval of 'minimal' episodic traces. Remembering and imagining may both be constructive, but they are so in fundamentally distinct ways (Robins 2022).

At its heart, then, the (dis)continuist debate concerns the operations of an episodic system. Are there specialised processes, selectively involved in memory tasks, that are distinctively *mnemic* in character? Even if there is a system or 'core network' active across tasks such as remembering one's past, imagining one's future, mind-reading, narrative comprehension, and event perception (cf. De Brigard (2025)), are there, e.g., selective retrieval processes of event-specific traces from an episodic store in the execution of only tasks of the first sort? This sort of question remains open and will be a locus of investigation for some time. The next 'wave' in memory science to shape philosophical theorising may emerge from the promise of engram neuroscience. (For recent discussions, see Andonovski et al. (2024), Camillo (2025), Najenson (2021), and Robins (2023).)

5 Current themes in philosophy of memory

In the current produced by the work surveyed in §4, philosophy of memory today is a vibrant and active sub-field. Recent topics and trends include going beyond systems-centric theorising to consider episodic-semantic interactions (Aronowitz 2025), the character of memory traces (De Brigard 2025), the accuracy conditions of remembering (Sant'Anna and Camillo, 2025), the function of episodic memory (Mahr 2022), the role of the social (McCarroll & Andonovski 2025), memory in nonhuman animals (Boyle & Brown 2025),

and aphantasia (Phillips 2025). Due to space constraints, here I will provide an introductory overview of two related topics: mnemic reference and memory errors.[8]

5.1 Mnemic reference

There is a sense of remembering in which it affords a means of singling out particulars—events, objects, places, individuals—in virtue of one's prior experience(s). When Sally remembers eating snake fruit for the first time last year, bringing to mind its distinctive taste and texture, she is directed at a specific event in her personal past. To emphasise this facet of remembering prompts the question: what, in any given case, determines of some particular that *it* is what one remembers? This question concerns the metasemantics of mnemic reference.

The first of two classic approaches is *descriptivism:* what one remembers is the satisfier of an associated cluster of descriptions. For instance, Sally's recollection involves property attributions and imagery one could construe as uniquely best satisfied by her first snake fruit experience. Yet pure descriptivist approaches face well-known challenges. For one, they rule out the possibility of remembering an event *e* for which one is unable to retrieve sufficiently many uniquely identifying descriptions. But if one's recollection was formed largely from information originating in experience of *e*, that should move the needle, in principle allowing mnemic reference to *e*.

According to *causalism*, by contrast, if one remembers an event *e* then one's experience of *e* left a memory trace, retained in the interim, the activation of which causally explains one's current representation (Martin & Deutscher 1966). The classical causalist's traces are *aetiologically discriminating,* embodying a link to the specific event during experience of which they were formed (Werning & Liefke 2025). Despite their elegance, purely causal approaches face challenges. First, it remains an open question whether there is an episodic system with a proprietary store of event-specific information eligible to play the 'memory trace' role (Langland-Hassan 2022b). Second, it may be difficult to extend the causal approach beyond event-specific remembering to cases where what we remember is an extended period, an object, a place, an individual, and so on.

[8] See Sant'Anna and Craver (Forthcoming) for a collection of introductions to contemporary work.

Recent approaches try to capture advantages of pure descriptivism/causalism while avoiding their costs. One such *hybrid* approach is reliabilist in character (Openshaw & Michaelian 2024): a memory representation refers to some particular (event, object, place) *x* just in case, given the psychological basis of its construction, *x* is the thing safely tracked across close possible worlds. So, for a contextually relevant range of features, the representation tends to attribute such a feature only if *x* exemplified it. The key emphasis here is less on providing an account that avoids any implicit talk of causation whatsoever, more on providing a characterisation that extends beyond specific events and allows more than only event-specific traces (e.g., schemata) to play a reference-fixing role. And while it is convenient to frame it as a biconditional analysis, its intended role is to encourage philosophers to broaden their sense of how mnemic reference-fixing might work. It remains to be seen whether some such approach can overthrow the causalist hegemony.

5.2 Memory errors

Memory errors can be roughly classified as omissive or commissive, depending on whether content that should be available is not or whether false/irrelevant content is produced. As hinted earlier (§4.1), our capacity for both is important. *Forgetting*—the loss of, or inability to retrieve, information—typically promotes efficiency. Our capacity for commissive errors is also often a product of otherwise beneficial operations, e.g., *misremembering* in the DRM paradigm. Of course, there are grave forms of each: amnesia and *confabulation*.

The term 'confabulation' originates in the work of 19th Century neurologists such as Korsakoff. He observed patients who would sincerely produce stories about their past that were (sometimes patently) false or markedly distorted. This 'pseudo-reminiscence' was often 'rooted in real memories' but somehow temporally displaced. The phenomenon is observed in patients with conditions such as Korsakoff's syndrome and Alzheimer's disease. For instance, one patient claimed he had married his wife only four months prior, rationalising their four adult children by claiming they were adopted (Moscovitch 1995).

Recent philosophy of memory has referred to confabulation, variously, as a type of error, a symptom, a disorder, and a malfunction. On a prominent view, they are wholly inaccurate apparent memories, "reflecting no influence of information retained from a particular past event" (Robins 2019: 2148). This has been a natural picture for causalists. The presence of an appropriate causal relation serves to secure mnemic reference *and*

separates genuine (mis)remembering from mere confabulation. Confabulation is, in key part, a distinctive form of *content error:* one involving reference-failure.

But are there cases of confabulation that involve mnemic reference (Openshaw 2025)? If there are, the causalist may be unable to explain them. Dominant *strategic retrieval* theories of confabulation provide a model of how this might occur. The offending breakdown may be nothing other than failure to retrieve a *task-relevant* trace, and to monitor success/failure in doing so (Burgess and Shallice 1996). Dalla Barba et al.'s (1990) Korsakoff's patient, C.A., frequently produced confabulations suggesting access to episodic traces (553) but that were temporally distorted. Asked what she did last Christmas, she claimed to have helped her mother cook lunch for her and her brothers, an event thought to have taken place, albeit before one of her brothers had died 30 years prior.

If we ought to sometimes credit confabulators with successfully mnemically referring to events in their past without crediting them with remembering, the simple causalist model is threatened. Confabulations would not be well-conceived as referentially empty 'mnemic hallucinations'. However, there is a natural line for causalists to take that would illustrate something positive and important about remembering. If confabulation may be attributable solely to malfunction of strategic retrieval (failure to identify a task-relevant episodic trace) and monitoring (failure to detect the first failure), it follows that there is no genuine remembering without properly functioning executive and monitoring processes. Entertaining an apparent memory representation that is appropriately causally connected to some event(s) in one's personal past may suffice for mnemic reference, but it cannot, on its own, suffice for genuine remembering. Albeit an appealing line, causalists will need to say more about the role of executive and monitoring processes within their architecture of episodic memory as a natural kind.

6 Conclusion

To tie up one loose thread, the exact relation between the scientific construct *episodic memory* and the putative subject matter of §3 is a matter of some debate. Some will view episodic memory as a successor concept and treat old debates as confused, historical curiosities. Others may take a more conservative approach, perhaps seeing in aspects of traditional debates a reflection of the 'manifest image' worth somehow saving. And others still may treat episodic memory as a failed natural kind, itself in need of replacement (§4.3).

At the same time, the methodological shift described in §4 has seen some traditional ideas fall on hard times: that there is a constitutively imagistic kind of remembering has perhaps fared less well than approximations of the previous awareness condition. Whether a different sort of life can be found in what seemed manifestly obvious will emerge as theoretical and metatheoretical developments continue to evolve and interact.

This has been a highly selective tour of a rich and rapidly developing field. Much of importance has been set aside. And there is no shortage of entry points for further inquiry, both across philosophical domains—mind, epistemology, ethics, and philosophy of science—and in interdisciplinary settings. Among the key issues expected to shape many questions in the years ahead are whether episodic memory is constituted by a mechanism distinct from episodic forms of imagining, what engram neuroscience is able to glean about memory traces, and how distinct kinds of memory interact. It is an exciting time for philosophy of memory, and fruitful dialogue with the sciences shows no sign of slowing.*

Works cited

Andonovski, N., Michaelian, K., 2024. Naturalism and simulationism in the philosophy of memory, in: Khani, A.H., Kemp, G., Amiriara, H., Rezaee, H.S. (Eds.), Naturalism and Its Challenges. Routledge, New York.

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

Anscombe, G.E.M. (Ed.), 1981. Memory, "experience", and causation, in: The Collected Philosophical Papers of G. E. M. Anscombe, Vol. 2: Metaphysics and the Philosophy of Mind. Blackwell, Oxford, pp. 120–130.

Aronowitz, S., 2025. Semanticization challenges the episodic–semantic distinction. The British Journal for the Philosophy of Science 76, 805–822. https://doi.org/10.1086/721760

* **Acknowledgements:** Many thanks to André Sant'Anna and Lauren Stephens for comments.
Funding: This project received funding from a Nanyang Technological University Start-Up Grant and was also supported by the French National Research Agency in the framework of the 'Investissements d'avenir' program (ANR-15-IDEX-02).

Aronowitz, S., 2018. Memory is a modeling system. Mind and Language 34, 483–502. https://doi.org/10.1111/mila.12220

Augustine, 1795. St. Augustine's Confessions: Or, Praises of God. P. Wogan, Dublin.

Ayer, A.J., 1956. The Problem of Knowledge. Harmondsworth, New York.

Baluška, F., Levin, M., 2016. On having no head: Cognition throughout biological systems. Front. Psychol. 7. https://doi.org/10.3389/fpsyg.2016.00902

Barkasi, M., Sant'Anna, A., 2022. Reviving the naïve realist approach to memory. Philosophy and the Mind Sciences 3. https://doi.org/10.33735/phimisci.2022.9192

Bartlett, F.C., 1932. Remembering: A Study in Experimental and Social Psychology. Cambridge University Press, Cambridge.

Bernecker, S., 2008. The Metaphysics of Memory. Springer Netherlands, Dordrecht. https://doi.org/10.1007/978-1-4020-8220-7

Boyle, A., Brown, S.A.B., 2025. Why might animals remember? A functional framework for episodic memory research in comparative psychology. Learn Behav 53, 14–30. https://doi.org/10.3758/s13420-024-00645-0

Brewer, W.F., 1996. What is recollective memory?, in: Rubin, D.C. (Ed.), Remembering Our Past: Studies in Autobiographical Memory. Cambridge University Press, Cambridge, pp. 19–66. https://doi.org/10.1017/CBO9780511527913.002

Broad, C.D., 1925. The Mind and its Place in Nature. Kegan Paul, Trench, Trubner & Co., London.

Buñuel, L., Israel, A., 2013. My Last Sigh. Vintage Books, New York.

Burgess, P.W., Shallice, T., 1996. Confabulation and the control of recollection. Memory 4, 359–411. https://doi.org/10.1080/096582196388906

Camillo, J.C., 2025. Interpreting continuism as a mechanistic thesis. Synthese 205, 140. https://doi.org/10.1007/s11229-025-04989-6

Cheng, S., Werning, M., 2016. What is episodic memory if it is a natural kind? Synthese 193, 1345–1385. https://doi.org/10.1007/s11229-014-0628-6

Colaço, D., 2022. What counts as a memory? Definitions, hypotheses, and "kinding in progress." Philosophy of Science 89, 89–106. https://doi.org/10.1017/psa.2021.14

Dalla Barba, G., Cipolotti, L., Denes, G., 1990. Autobiographical memory loss and confabulation in Korsakoff's syndrome: a case report. Cortex 26, 525–534. https://doi.org/10.1016/s0010-9452(13)80302-4

De Brigard, F., 2025. Simulationism and Memory Traces, in: Nadel, L., Aronowitz, S. (Eds.), Space, Time, and Memory. Oxford University Press.

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

Debus, D., 2008. Experiencing the past: A relational account of recollective memory. Dialectica 62, 405–432. https://doi.org/10.1111/j.1746-8361.2008.01165.x

Ebbinghaus, H., 1885. Über das Gedächtnis: Untersuchungen zur experimentellen Psychologie. Duncker & Humblot, Leipzig.

Feest, U., 2025. Operationism in Psychology: An Epistemology of Exploration. University of Chicago Press, Chicago, IL.

Furlong, E.J., 1951. A Study in Memory: A Philosophical Essay. Thomas Nelson & Sons, London.

Goldwasser, S., 2023. Memory as skill. Rev.Phil.Psych. 14, 833–856. https://doi.org/10.1007/s13164-021-00605-x

Hassabis, D., Kumaran, D., Vann, S.D., Maguire, E.A., 2007. Patients with hippocampal amnesia cannot imagine new experiences. Proceedings of the National Academy of Sciences 104, 1726–1731. https://doi.org/10.1073/pnas.0610561104

Hobbes, T., 2011. Leviathan, Revised student edition, 14th printing. ed. Cambridge University Press, Cambridge.

Hoerl, C., 2022. A knowledge-first approach to episodic memory. Synthese 200, 1–27. https://doi.org/10.1007/s11229-022-03702-1

Hoerl, C., 2018. Remembering past experiences: Episodic memory, semantic memory and the epistemic asymmetry, in: Michaelian, K., Debus, D., Perrin, D. (Eds.), New Directions in the Philosophy of Memory. Routledge, New York, pp. 313–328.

Hopkins, R., 2024. The Profile of Imagining. Oxford University Press, Oxford.

Hopkins, R., 2018. Imagining the past: On the nature of episodic memory, in: Dorsch, F.M.F. (Ed.), Memory and Imagination. Oxford University Press.

Hume, D., 1978. A Treatise of Human Nature, Second edition, P. H. Nidditch (Ed.). ed. Oxford University Press, Oxford.

Klein, S.B., 2015. What memory is. WIREs Cognitive Science 6, 1–38. https://doi.org/10.1002/wcs.1333

Langland-Hassan, P., 2022a. Remembering and imagining: The attitudinal continuity, in: Berninger, A., Ferran, I.V. (Eds.), Philosophical Perspectives on Memory and Imagination. Routledge.

Langland-Hassan, P., 2022b. Remembering, imagining, and memory traces: Toward a continuist causal theory, in: Sant'Anna, A., McCarroll, C., Michaelian, K. (Eds.), Current Controversies in Philosophy of Memory. Current Controversies in Philosophy.

Levine, B., Svoboda, E., Hay, J.F., Winocur, G., Moscovitch, M., 2002. Aging and autobiographical memory: Dissociating episodic from semantic retrieval. Psychology and Aging 17, 677–689. https://doi.org/10.1037/0882-7974.17.4.677

Liefke, K., Werning, M., 2024. Diachronicity matters! How semantics supports discontinuism about remembering and imagining. Topoi 43, 1137–1159. https://doi.org/10.1007/s11245-024-10068-1

Locke, D., 1971. Memory. Macmillan, London.

Locke, J., 1690. An Essay Concerning Human Understanding. Oxford University Press, Oxford.

Loftus, E.F., Miller, D.G., Burns, H.J., 1978. Semantic integration of verbal information into a visual memory. Journal of Experimental Psychology: Human Learning and Memory 4, 19–31. https://doi.org/10.1037/0278-7393.4.1.19

Mahr, J.B., 2022. Episodic memory: And what is it for?, in: Sant'Anna, A., McCarroll, C., Michaelian, K. (Eds.), Current Controversies in Philosophy of Memory. Current Controversies in Philosophy.

Malcolm, N. 1963. *Knowledge and Certainty: Essays and Lectures*. Prentice-Hall, Englewood Cliffs, NJ.

Martin, C.B., Deutscher, M., 1966. Remembering. Philosophical Review 75, 161–196. https://doi.org/10.2307/2183082

Martin, M.G.F., 2019. Betwixt feeling and thinking: Two-level accounts of experience, in: Knowles, J., Raleigh, T. (Eds.), Acquaintance. Oxford University Press, Oxford, pp. 95–126. https://doi.org/10.1093/oso/9780198803461.003.0005

Martin, M.G.F., 2015. Old acquaintance: Russell, memory and problems with acquaintance. Analytic Philosophy 56, 1–44. https://doi.org/10.1111/phib.12059

McCarroll, C.J., 2022. Memory and imagination, minds and worlds, in: Philosophical Perspectives on Memory and Imagination. Routledge.

McCarroll, C.J., Andonovski, N., 2025. Episodic memory as a mindshaped capacity, in: The Routledge Handbook of Mindshaping. Routledge.

Michaelian, K., 2016. Mental Time Travel: Episodic Memory and Our Knowledge of the Personal Past. MIT Press, Cambridge, MA.

Michaelian, K., 2011. Generative memory. Philosophical Psychology 24, 323–342. https://doi.org/10.1080/09515089.2011.559623

Michaelian, K., Sutton, J. 2017. Memory, in: The Stanford Encyclopedia of Philosophy. https://plato.stanford.edu/archives/sum2017/entries/memory/

Moscovitch, M., 1995. Confabulation, in: Memory Distortions: How Minds, Brains, and Societies Reconstruct the Past. Harvard University Press, Cambridge, MA, pp. 226–251.

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

Openshaw, J., 2025. Illusions of memory: What referential confabulation can tell us about remembering. Asian J. Philos. 4, 83. https://doi.org/10.1007/s44204-025-00305-8

Openshaw, J., 2022. Remembering objects. Philosophers' Imprint 22. https://doi.org/10.3998/phimp.699

Openshaw, J., Michaelian, K., 2024. Reference in remembering: towards a simulationist account. Synthese 203, 90. https://doi.org/10.1007/s11229-024-04508-z

Perrin, D., 2016. Asymmetries in subjective time, in: Seeing the Future: Theoretical Perspectives on Future-Oriented Mental Time Travel. Oxford University Press, New York, NY, US, pp. 39–61. https://doi.org/10.1093/acprof:oso/9780190241537.003.0003

Phillips, I., 2025. Aphantasia reimagined. Noûs n/a. https://doi.org/10.1111/nous.12551

Reid, T., 1785. Essays on the intellectual powers of man. John Bell, Edinburgh.

Reid, T., 1765. An inquiry into the human mind, on the principles of common sense, Second edition. ed. Kincaid & Bell, Edinburgh.

Robins, S., 2023. The 21st century engram. Wiley Interdiscip Rev Cogn Sci 14, e1653. https://doi.org/10.1002/wcs.1653

Robins, S.K., 2022. Episodic memory is not for the future, in: Sant'Anna, A., McCarroll, C., Michaelian, K. (Eds.), Current Controversies in Philosophy of Memory. Routledge, New York, pp. 166–184.

Robins, S.K., 2019. Confabulation and constructive memory. Synthese 196, 2135–2151. https://doi.org/10.1007/s11229-017-1315-1

Roediger, H.L., McDermott, K.B., 1995. Creating false memories: Remembering words not presented in lists. Journal of Experimental Psychology: Learning, Memory, and Cognition 21, 803–814. https://doi.org/10.1037/0278-7393.21.4.803

Russell, B., 2005. The Analysis of Mind. Routledge, London. https://doi.org/10.4324/9780203996515

Sant'Anna, A., 2023. Is remembering constructive imagining? Synthese 202, 141. https://doi.org/10.1007/s11229-023-04338-5

Sant'Anna, A., Camillo, J.C., 2025. The debate about the accuracy conditions of episodic memory: a critical overview. Asian J. Philos. 4, 120. https://doi.org/10.1007/s44204-025-00347-y

Sant'Anna, A., Craver, C.F. (Eds.), Forthcoming. The Oxford Handbook of Philosophy of Memory. Oxford University Press.

Schacter, D.L., Addis, D.R., 2007. The cognitive neuroscience of constructive memory: remembering the past and imagining the future. Philos Trans R Soc Lond B Biol Sci 362, 773–786. https://doi.org/10.1098/rstb.2007.2087

Schacter, D.L., Tulving, E., 1994. What are the memory systems of 1994?, in: Memory Systems 1994. MIT Press, Cambridge, MA, pp. 1–38.

Sellars, W. 1997. Empiricism and the Philosophy of Mind. Harvard University Press, Cambridge, MA.

Sorabji, R., 1972. Aristotle on Memory, Second Edition, 2004. ed. Bloomsbury.

Soteriou, M., 2013. The Mind's Construction: The Ontology of Mind and Mental Action. Oxford University Press, Oxford, United Kingdom.

Squire, L.R., 2004. Memory systems of the brain: a brief history and current perspective. Neurobiol Learn Mem 82, 171–177. https://doi.org/10.1016/j.nlm.2004.06.005

Tooming, U., Miyazono, K., 2024. Prospects for epistemic generationism about memory. PhiMiSci 5. https://doi.org/10.33735/phimisci.2024.10248

Tulving, E., 1991. Concepts of human memory, in: Memory: Organization and Locus of Change. Oxford University Press, Oxford, pp. 3–32.

Tulving, E., 1985. Memory and consciousness. Canadian Psychology / Psychologie canadienne 26, 1–12. https://doi.org/10.1037/h0080017

Tulving, E. 2002. Episodic memory: From mind to brain. Annual Review of Psychology 53. 10.1146/annurev.psych.53.100901.135114

Wearing, D., 2005. Forever Today: A Memoir of Love and Amnesia. Corgi.

Werning, M. 2020. Predicting the past from minimal traces: Episodic memory and its distinction from imagination and preservation. Review of Philosophy and Psychology 11(2). https://doi.org/10.1007/s13164-020-00471-z

Werning, M., Liefke, K., 2025. Reference in memories from perceptual and non-perceptual experiences: a non-disjunctivist account including vicarious, oneiric and fictional remembering. Synthese 205, 186. https://doi.org/10.1007/s11229-025-04936-5

Zemach, E.M., 1968. A definition of memory. Mind 77, 526–536. https://doi.org/10.1093/mind/lxxvii.308.526