

1 Three Questions about Memory

The argument of this book is driven by three core questions:

- What is memory?
- How does memory give us knowledge?
- When and why did memory emerge?

This introductory chapter first provides a brief review of the philosophical and psychological relevance of each of these questions and then sketches the general plan of the book's argument.

1.1 What Is Memory?

The attempt to say what memory *is* amounts, in the first place, to an attempt to give an informative general characterization of remembering—of what it is for someone to remember something. When contemporary philosophers pursue this project, they typically do so by formulating, more or less explicitly, sets of conditions meant to be individually necessary and jointly sufficient for the truth of statements of the form *S remembers x*. Though the book makes use of this sort of analytic methodology in places, it does so primarily as a matter of convenience. Ultimately, the goal is not to develop a theory that can necessarily be adequately expressed in the form of an analysis, but rather to formulate a useful general framework for thinking about human memory, one that draws out and makes explicit the vision of memory implicit in current psychology and that can in turn contribute to interpreting the findings of and suggesting new lines of inquiry for the latter. Consequently, the reasoning employed here is not confined to the a priori “theory → counterexample → revised theory” style of argumentation associated with *S remembers x*-style theories. The goal, in short, is not

to develop an account of remembering immune to all possible counterexamples, no matter how far out, but rather to develop an account that captures what is important about *human* remembering, in particular, as it unfolds in the real world.

Interest in the nature of remembering goes back essentially to the dawn of philosophy, and most major philosophers have had something to say about it (Burnham 1888). This is not the place for a comprehensive historical review—an undertaking which would require a book unto itself—but a few highlights will help to orient us. Early on, we find Aristotle remarking that

one cannot remember the future, but of this one has opinion and expectation ... ; nor can one remember the present, but of this there is sensation; for by sensation we cognise neither the future nor the past but only the present. ... Memory is of the past. (Bloch 2007, 25)

Jumping forward in time, we have Hume (1739) arguing for his well-known vivacity criterion for memory, according to which memory has a distinctive kind of content. More recently, Martin and Deutscher (1966) have influentially argued that memory requires a specific kind of causal connection with the past. All three of these views are intuitively plausible, and all have been widely taken up. All three—the view that memory is of the past, the view that memory and imagination are distinguished by their respective contents, and the view that memory necessarily involves a causal connection with the past—are challenged in this book, and, if its argument succeeds, none survives unscathed.

As the examples of Aristotle, Hume, and Martin and Deutscher suggest, what has driven philosophers to develop general theories of remembering is ultimately a question about demarcating memory from related cognitive phenomena: in virtue of what is it the case that a subject on a given occasion counts as remembering *rather than* undergoing some other cognitive process? Aristotle, for example, is concerned with the difference between remembering, on the one hand, and perception and “expectation,” on the other, where expectation coincides roughly with what we refer to today as future-oriented mental time travel (Klein 2013a). Typically, the concern has been with the difference between remembering and imagining, where the latter is construed sufficiently broadly to include a range of forms of imagination in addition to future-oriented mental time travel. Hume’s vivacity criterion is intended, on one reading, to capture the difference

between remembering events and imagining them. And Martin and Deutscher's causal condition is in part a rival attempt to capture that same difference.

Remembering, like imagining (Kind 2013), is a multifarious phenomenon, but this book is concerned specifically with the form of memory that philosophers have referred to as (simply) memory (Earle 1956), memory par excellence (Bergson 1896), true memory (Russell 1921), perceptual memory (Broad 1925; Malcolm 1963), reminiscence (Ryle 1949), retrospective memory (Furlong 1948), event memory (Ayer 1956), recollection (von Leyden 1961), or personal memory (Locke 1971)—to repeat just part of a list compiled by Brewer (1996). In psychology, Tulving has influentially dubbed this form of memory *episodic* (Tulving 1972, 1983), and the term is increasingly catching on in philosophy (Perrin and Rousset 2014). Episodic memory refers, roughly, to the form of memory responsible for allowing us to revisit specific episodes or events from the personal past. It is typically contrasted with semantic memory, which allows us to recall facts without necessarily giving us access to the episodes in which they were learned. Given that we are interested in distinguishing remembering from imagining, and given that we are interested specifically in episodic memory, the question for the theory of remembering becomes: What distinguishes episodic remembering from episodic imagining?

The distinction between episodic imagination and semantic imagination has been referred to in a variety of ways, including imagination versus supposition (Szabó Gendler 2010), perceptual versus propositional imagination (Currie and Ravenscroft 2002), and enactive imagination versus supposition (Goldman 2006). Semantic imagination is, roughly, the form of imagination responsible for our ability to suppose that something is the case (see Byrne 2007); it is analogous to semantic memory, memory for facts. Episodic imagination, in contrast, is the form of imagination responsible for our ability to entertain possible episodes, just as episodic memory is responsible for the ability to revisit past episodes. The distinction between episodic and semantic imagination may be a matter of degrees (Szpunar et al. 2014); imagining generic events, for example, may draw on both episodic and semantic imagination. But our focus here will be on purely episodic forms of imagination.

The question of how remembering is distinguished from imagining derives some of its philosophical relevance from the question of personal

identity. Since Locke (1689), episodic memory has figured centrally in accounts of what makes someone count as the same person over time.¹ While personal identity as such is beyond the scope of this book, there is a loose correspondence between the Lockean view that personal identity is determined by memory and the view current in psychology that memory is constitutive of the self (Cosentino 2011; Rubin 1999). More importantly, we will see that the sense of self turns out to be crucial to the ability to reliably remember the past.

The question of the nature of memory derives additional philosophical relevance from its link to the question of the nature of memory knowledge. It is intuitively plausible that, in order for memory to function as a source of knowledge, there must be a difference in kind between remembering and imagining. While the book will argue that memory can in fact provide knowledge of the past even if there is no such difference, the question of the objective difference between memory—viewed as one form of imagination—and *other* forms of imagination continues to be of crucial importance to epistemology, since it is intimately related to the subjective difference or differences on which agents rely to determine whether they are remembering or imagining; hence the question of the nature of remembering remains a necessary prelude to the epistemology of memory.

The philosophical importance of an account of the nature of remembering is relatively uncontroversial. Are philosophical theories of memory likewise relevant to psychology? The answer may depend, to some extent, on the kind of theory at issue, as some philosophical theories are couched in terms that get little empirical traction. This goes, fairly clearly, for the epistemic theory of memory, according to which memory is, roughly, retained knowledge (see, e.g., Holland 1974; Munsat 1967; Zemach 1968). Arguably, it goes as well for recent accounts such as that defended by Martin (2001), according to which remembering is a matter of retaining “acquaintance,” or that defended by Byrne (2010), according to which remembering is a matter of preserving “cognitive contact.” Such theories, at any rate, will not be discussed in any detail here.

In contrast, other philosophical theories of remembering—including the empiricist theory, the causal theory, and the simulation theory defended in this book—can indeed be brought into contact with the relevant psychology, though it may in some cases take some work to do so. In fact, the simulation theory—according to which there is no essential difference

between remembering and imagining, with memory turning out to be simply one form of imagination among others—is intended in part to be a generalization and synthesis of a cluster of views on the nature and function of memory that have emerged recently in psychology in response to the growing recognition that episodic memory is intimately related to a broad range of episodic constructive processes, including future-oriented mental time travel, episodic counterfactual thinking, and mindreading (Szpunar 2010).

Psychologists may, of course, investigate memory without first having fully defined it (Feest 2011), but, as our understanding of memory deepens, it is appropriate to periodically take stock, as the findings that have accumulated may force us to reevaluate our intuitive understanding of memory. The simulation theory is meant to contribute to such a reevaluation. The theory also aims, more broadly, to contribute to reshaping the metaphors that guide both everyday and scientific thinking about memory. Current popular thinking about memory is largely dominated by the metaphor of computer memory (Michaelian 2012). As Draaisma (2000) has shown, while it is natural for us today to conceive of human memory on the model of computer memory, there is, in light of the extraordinary variety of metaphors that have historically been used to describe memory, little reason to suppose that this particular metaphor will be the last. Indeed, in one respect, the computer metaphor is not even particularly novel: human memory has often been compared to new technologies, in particular to forms of external memory or information storage. Thus Plato (in the *Theaetetus*) compared memory to a wax tablet, and it has since been said to be analogous to libraries, archives, photographs, films, holograms, and so on.

These metaphors represent remembering as being essentially a matter of storing and retrieving fixed items of information. But Draaisma reviews another family of memory metaphors, many of them organic rather than technological, including woods, fields, labyrinths, caves, grottoes, mine-shafts, the depths of the sea, palaces, abbeys, and theaters, which represent remembering as a complex, effortful, and failure-prone activity. The existence of this second set of metaphors suggests a recognition of the limited utility of a purely storage-and-retrieval conception and the need for a conception which emphasizes remembering as a process in which the agent actively reconstructs the past. Both of these competing conceptions

are alive in contemporary psychology, tracing back to foundational work by Ebbinghaus (1885) and Bartlett (1932), respectively.

Ebbinghaus's work, leading to his famous forgetting curve (describing the way in which successful retrieval reduces exponentially as a function of time), focused on the capacity to remember lists of nonsense syllables. This quantity-oriented conception of memory, which views memory as a sort of storehouse into which arbitrary contents can be placed, was crucial to the establishment of experimental psychology and dominated memory research long after Ebbinghaus, for example in the form of the standard list-learning paradigm. In recent years, however, the rival accuracy-oriented conception (Koriat and Goldsmith 1996), which focuses on the correspondence (or noncorrespondence) between memory and reality, has become increasingly influential, due in part to the increasing salience of memory errors and distortions (Brainerd and Reyna 2005). Bartlett's own research focused in part on the influence of schemas in shaping retrieved content, and similar themes have been prominent in recent memory research, for example, in research on the effects of conversation on what is remembered (Stone et al. 2009).

The theory developed here, with its focus on mental time travel as a simulational process, is meant to reinforce the accuracy-oriented conception of memory. Realistically, however, the tension between quantity-oriented and accuracy-oriented approaches is unlikely to be resolved in the near future, for, while it is difficult to incorporate both views within a single model, each captures an important aspect of the workings of memory. Remembering *is* thoroughly constructive, but the simulation theory, in contrast to some Wittgensteinian views (Moyal-Sharrock 2009; Stern 1991), does assign an important role to information storage. Nevertheless, given the current predominance of the computer metaphor in popular discussions of memory, it may be useful to tip the scales in the other direction.

1.2 How Does Memory Give Us Knowledge?

Memory is sometimes listed as a potential "basic" epistemic source alongside other candidate basic sources, including perception, introspection, inference, and communication. There is no uncontroversial list of basic epistemic sources, however, and it is unclear whether memory is basic in an interestingly strict sense.

Relying on Audi's criterion of epistemic basicness, for example, a source is basic if it yields knowledge without dependence on the operation of some other source (Audi 2002). Given that it depends on input from other sources, memory straightforwardly fails to satisfy this criterion. In fact, however, it is unclear whether *any* source satisfies Audi's criterion. Consider perception, our best candidate for a basic epistemic source. Perception appears to be cognitively penetrable, in the sense that top-down expectations influence visual experience (Vetter and Newen 2014). There is increasing evidence for a multisource model of visual perception, according to which visual input provides only part of the content of visual perception, with perceptual representations in effect being simulations driven by a number of factors, including stored information (Intraub 2010). If this model is on the right track, perception is no more independent of memory than memory is of perception. The criterion may thus simply be too strong to be useful, and memory may satisfy a suitably weakened criterion.

But regardless of whether memory is, strictly speaking, epistemically basic, it is uncontroversial that it is basic in the loose sense of being a *core* epistemic source: take memory away, and much—perhaps virtually all—human knowledge goes along with it, including an apparently distinctively human form of knowledge. As many epistemologists have argued, at any given time, most of our beliefs are not occurrent but merely dispositional—that is, on a natural conception of dispositional belief, they are stored in memory.² Moreover, an important—and probably distinctively human (Corballis 2011)—form of knowledge depends on memory in an especially strong sense. It is possible for a subject to gain knowledge of his own past without that knowledge amounting to memory knowledge; for example, one typically learns about one's early childhood through testimony from one's family. But a central form of knowledge of one's past—the sort of knowledge that you can have of your own past but not of anyone else's—depends essentially on (episodic) memory.

There is thus a need for an account of how memory functions as an epistemic source and, in particular, of how it functions to give us knowledge of the personal past. Moreover, once the constructive character of memory is taken into account, the threat of skepticism about memory knowledge looms. This is not Russell-style skepticism, based on the observation that, given that there is no logically or metaphysically necessary connection between an apparent memory and the occurrence of the relevant event, the

hypothesis that the world came into existence five minutes ago, complete with our apparent memories, cannot be ruled out (Russell 1921). The threat is, rather, from an empirically grounded form of skepticism. Given the constructive character of remembering, opportunities for error and distortion abound. In light of this, it is far from obvious that memory can meet plausible standards for knowledge—not the impossibly high standards of the traditional skeptic but, as Shanton (2011) has pointed out, moderate, well-motivated standards, including reliability.

This book—the basic epistemological claim of which is that, due in part to the adaptive character of the constructive processes involved in remembering, and due in part to the sensitivity of the metacognitive monitoring processes involved in remembering, episodic memory avoids the threat of unreliability—is in part a response to this skeptical threat, with a focus on the standard of reliability. The second part of this epistemological claim, according to which metacognition (the monitoring and control of mental processes; Dunlosky and Metcalfe 2008; Proust 2010) enables remembering subjects to compensate for the uncertainty introduced by the constructive character of remembering in order to minimize unreliability, requires a fair bit of background before it can be meaningfully unpacked. The first part of the claim is somewhat more straightforward. As Schacter once put it, with memorable understatement, “the output of human memory often differs—sometimes rather substantially—from the input” (1995). The claim defended here is that the principles governing constructive processing in remembering tend to result in “retrieved” memories that are accurate despite departing from the relevant stored information, which itself may depart from the information initially encoded, which in turn may depart from the information contained in the original experience of the relevant episode.

One version of the skeptical worry is based on research on the way in which a subject’s memories can be shaped and reshaped by interaction with other agents. The focus of the research in question is not on direct interventions in memory (via, for example, pharmacological agents; see Liao and Sandberg 2008) but rather on the routine incorporation, without the agent’s awareness, of communicated information into the apparently purely experiential representations produced by reconstructive memory retrieval. Work by Loftus (1996) and others on this sort of incorporation in the context of eyewitness memory led to the notion of the “misinformation effect,” in

which the incorporation of testimonial information results in inaccurate (or less accurate) memory for experienced events. It might seem obvious that such incorporation can only decrease the accuracy of memory. But the typical experimental setup in this area focuses precisely on providing subjects with misleading post-event information. If it turns out that, in ecological settings, subjects are more likely to receive accurate than inaccurate post-event information, then the incorporation of testimonial information may actually increase the accuracy of retrieved memories, despite resulting in retrieved memories that depart from the experience of events.

The suggestion is, in effect, that there is a sort of “preestablished harmony” between the way in which memory tends to simulate episodes and the shape of the episodes that we actually experience. This suggestion deliberately echoes Kornblith’s claim that there is a preestablished harmony between the way in which we tend to perform inductive inferences, which departs dramatically from the canons of good statistical reasoning, and the way in which the properties we project in our inductive inferences are actually grouped in the world (Kornblith 1995). A priori, our tendency to project properties on the basis of extremely small samples—Tversky and Kahneman’s “law of small numbers” (1971)—would appear to bode ill for the possibility of knowledge based on inductive inference. But it turns out, Kornblith argues, that this tendency aligns with relevant patterns of properties in the world in such a way that our inductive inferences are nevertheless reliable. Analogously, the constructive, simulational character of our remembering, as opposed to the sort of purely reproductive, computer-style memory that intuition tells us would be preferable, would seem to be bad news for the possibility of memory knowledge. But it turns out, according to the simulation theory, that the relevant constructive processes are engineered in such a way that remembering is nevertheless reliable overall.

The psychological relevance of this account of memory knowledge derives primarily from its emphasis on the mechanisms responsible for ensuring the reliability of potentially unreliable constructive remembering. Such an account may serve as a corrective to the tendency sometimes displayed by psychologists to emphasize errors and distortions due to constructive processes to such an extent that the overall message is that remembering is outright *unreliable*. Critiquing this tendency, Ost and Costall make the reasoning explicit: “remembering is intrinsically reconstruction *and hence inevitably unreliable*” (2002, 246; emphasis added). As Kornblith

argues, the initial, pessimistic reaction to findings of apparent inductive irrationality was unwarranted: just as perceptual illusions and errors are studied not because they show that perception is unreliable but rather because they provide, by showing where the mechanisms responsible for the reliability of perception break down, a glimpse of the workings of those mechanisms, inferential errors should be studied because they provide a glimpse of the workings of the mechanisms responsible for the overall reliability of inductive inference. Likewise, the sometimes pessimistic tone of constructive memory research, or at least of commentary on such research, is unwarranted: memory errors and illusions, when all goes well, play a role in memory research analogous to that played by perceptual errors and illusions in perception research, providing a means of investigating the operation of the constructive processes which, under normal conditions, ensure the overall reliability of remembering (Roediger 1996).

1.3 When and Why Did Memory Emerge?

While much of this book is devoted to building a philosophical account of remembering and memory knowledge on the current psychology of constructive memory and mental time travel, the intention is ultimately for that account to feed back into psychology. Hence the book's third question, concerning the evolution of episodic memory, belongs more to psychology than it does to philosophy.

Not long ago, Boyer remarked that there was relatively little work in which a functional, evolutionary perspective on memory is adopted (Boyer 2009). While there were important early exceptions to this generalization (e.g., Anderson 1990; Anderson and Milson 1989), the situation has since changed dramatically, and there is now a fast-developing literature exploring the evolution of memory. The debate over the survival-processing paradigm, which suggests that memory is tuned to remember the kinds of scenarios that tended to be especially survival-relevant in ancestral environments, accounts for a significant fraction of this literature (Nairne and Pandeirada 2008; Schwartz et al. 2014), but it is the debate over the existence of mental time travel in nonhuman animals that is most directly relevant here (Cheke and Clayton 2010). That debate—essentially a debate over *when* episodic memory evolved—turns in part on the interpretation of experimental evidence, but it also turns on a conceptual or definitional

matter. If the subjective dimension of human episodic memory—its characteristic phenomenology—is essential to it, then the research that has demonstrated a form of episodic-like memory—memory for specific past episodes—in certain nonhuman species is insufficient to demonstrate that the latter are capable of full-blown episodic memory. According to the simulation theory, the subjective dimension is not, strictly speaking, essential to *memory*, but it is essential to its *reliability*. There is thus good reason to regard nonhuman episodic-like memory, the reliability of which does not depend on such a subjective dimension, as being different in kind from human episodic memory. This insight concerning the necessity of phenomenology for reliability grounds an intervention in the parallel debate over *why* episodic memory evolved. On the view defended here, a capacity for simulational remembering could not have evolved unaccompanied by a capacity for the relevant phenomenology.

1.4 Overview

Part I sets the stage for the rest of the book. Relying on research on the memory systems of the human brain, chapter 2 argues that the deep differences among different kinds of memory militate against any attempt to develop a unified approach to the epistemology of memory as a whole, and makes a case for starting with episodic memory in particular. Chapter 3 sets out the epistemological framework that guides the book's treatment of episodic memory; the framework combines a strong form of naturalism, which urges us to come to grips with empirical research on constructive memory and mental time travel, with a weak form of reliabilism, which suggests a focus on the effects of metacognitive monitoring on the reliability of remembering.

The simulation theory of memory is developed and defended in part II. Chapters 4 and 5 look at research on constructive memory, exploring the prospects for updating the causal theory of memory to enable it to accommodate the constructive character of remembering. Chapter 6 argues that mental time travel research shows that construction in memory is so extensive that it requires abandoning the causal condition entirely. The result is a simulation theory of memory. Whereas the causal theory claims that remembering and imagining are distinguished by the presence, in the case of the former, of an appropriate causal connection with the relevant past

episode, the simulation theory claims that there is no intrinsic difference between remembering and imagining—to remember, it turns out, is just to imagine the past.

Part III takes up the question of the reliability of simulational remembering. The goal of this part is not to show *that* remembering is reliable; as epistemologists have often observed, that is not something that could be shown in a noncircular fashion, since any demonstration of the reliability of memory would itself have to rely on memory. Instead, the account is meant to show *how* remembering might achieve reliability despite the opportunities for error and distortion that are inevitably entailed by its simulational character. The argument proceeds in two steps. The first step considers remembering in isolation from other forms of episodic imagination, looking at how it avoids being overwhelmed by error and distortion despite recombining and transformation information so extensively that a “retrieved” memory might end up including little to no information originating in experience of the relevant episode. The claim is that massive error and distortion are avoided due, first, to the reliable character of the constructive processes internal to memory (chapter 7) and, second, to the role of metacognitive monitoring, including source monitoring, in filtering out information originating in unreliable sources (chapter 8). The second step of the argument (chapter 9) deals with a complication that arises when forms of episodic imagination other than episodic memory are taken into account. Episodic remembering would be unreliable were we unable to distinguish it from other forms of episodic imagination. But there is extensive similarity—at the neural, cognitive, and phenomenological levels—between memory and other forms of imagination, raising the question of how we manage to avoid confusing different forms of imagination. The claim defended in this chapter is that agents avoid confusion due to their capacity for a form of metacognition, what we can refer to as “process monitoring,” in which the characteristic phenomenology of remembering—consciousness of the self in subjective time—plays an important role.

Part IV turns to the evolution of episodic memory. Chapter 10 surveys the main lines of the debate over the evolution of episodic memory, arguing that, relative to the forms of episodic-like memory that have been demonstrated in animals, human episodic memory has two distinctive features. First (as emphasized in part II), it is one function of a general episodic construction system capable of simulating a wide range of possible past and

future episodes. Second (as emphasized in part III), it involves a distinctive form of consciousness of the self in subjective time. Chapter 11 argues that the adaptivity of the first feature is relatively straightforward, and that the first feature explains the second. Existing explanations account for the adaptivity of a capacity to simulate a range of possible episodes, as opposed to simply retrieving information about what happened where and when, as in animal episodic-like memory, but they have a harder time accounting for the characteristic phenomenology of remembering. The claim defended in this chapter is that a general simulational capacity inevitably entails uncertainty about whether simulated episodes actually occurred and that the need for process monitoring suggests that the phenomenology of remembering plays a role in reducing this uncertainty to a manageable level. Finally, chapter 12 brings things to a close, briefly summing up the book's main conclusions.