which is a candidate mechanism for synaptic changes in memory formation. Although some kinds of LTP show parallels with some kinds of memory formation, it has not yet been established that any of the kinds of LTP is the mechanism for any of the kinds of memory formation.

The discussion above has been limited, for simplicity, to increases in activity and excitatory processes at synaptic junctions, but memory may equally well involve decreases in activity and inhibitory processes.

## CONCLUSION

The establishment of memory appears to employ a cascade of neurochemical events operating on both sides of the synaptic junction. In some cases this is complemented by structural changes in axon terminals, dendritic spines, and location of receptor molecules. It remains to be determined which of these changes may be necessary and sufficient for memory formation.

### Further Reading

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

Finger S (1994) *Origins of Neuroscience: A History of Explorations into Brain Function*. New York, NY: Oxford University Press.

Hayashi Y, Shi SH, Esteban JA *et al*. (2000) Driving AMPA receptors into synapses by LTP and CaMKII: requirement for GluR1 and PDZ domain interaction. *Science* **287**: 2262–2267.

Kandel ER (2000) Learning and memory. In: Kandel ER, Schwartz JM and Jessell TM (eds) *Principles of Neural Science*, 4th edn. New York, NY: McGraw-Hill.

Kety SS (1976) Biological concommitants of affective states and their possible role in memory processes. In: Rosenzweig MR and Bennett EL (eds) *Neural Mechanisms of Learning and Memory*, pp. 321–326. Cambridge, MA: MIT Press.

Martin SJ, Grimwood PD and Morris RGM (2000) Synaptic plasticity and memory: an evaluation of the hypothesis. *Annual Review of Neuroscience* **23**: 649–711.

Rosenzweig MR (1996) Aspects of the search for neural mechanisms of memory. In: *Annual Review of Psychology* **47**: 1–32. Palo Alto, CA: Annual Reviews.

Rosenzweig MR (1998) Historical perspectives on the development of the biology of learning and memory. In: Martinez JL and Kesner RP (eds) *Neurobiology of Learning and Memory*, 3rd edn, pp. 1–33. San Diego, CA: Academic Press.

Rosenzweig MR, Breedlove SM and Leiman AL (2001) *Biological Psychology: An Introduction to Behavioral, Cognitive, and Clinical Neuroscience*, 3rd edn. Sunderland, MA: Sinauer.

# Memory, Philosophical Issues about

Intermediate article

*John Sutton*, Macquarie University, Sydney, New South Wales, Australia

**CONTENTS**



*Memory is a set of cognitive capacities by which humans and other animals retain information and reconstruct past experiences, usually for present purposes. Philosophical investigation into memory is in part continuous with the development of cognitive scientific theories, but includes related inquiries into metaphysics and personal identity.*

## INTRODUCTION

Cognitive scientists study an enormous variety of topics under the headings of memory and learning. Events affect humans, other animals, and machines in many ways, and sometimes enduring changes in the respective systems result, altering actual and possible future behavior. Philosophers have been centrally interested in human personal memory for episodes and experiences in the autobiographical past, as manifested in reminiscence, recall, and recognition. This is partly because of a focus on memory as a topic in epistemology, the theory of knowledge, which was in the twentieth century often pursued through methods of linguistic and

conceptual analysis that were oddly divorced from psychological inquiry. Recently, however, a naturalistic turn in philosophy of mind has brought theories of memory firmly into productive contact with cognitive science.

This article first sketches central philosophical questions about memory, self, and time. It then focuses on a specific debate which is particularly relevant to cognitive science, about the existence and nature of memory traces, and concludes by outlining recent moves towards a more insistently interdisciplinary approach to memory.

## PHILOSOPHICAL ISSUES CONCERNING MEMORY

### Memory and Self

Cognitive scientific views of consciousness and self will increasingly influence traditional philosophical discussion about memory and personal identity. To what extent does memory construct and maintain the continuity of personal identity over time? In the history of Western philosophy this question was of pressing concern in religious contexts. In order for it to be truly *me* who is saved or damned at the Day of Judgment, the judged soul has to be numerically identical with the person who committed sinful or praiseworthy acts in this life. For Christian philosophers like John Locke, this meant that we had to retain personal memory in the afterlife.

In our 'materialist' age, the issue remains urgent because many moral and legal practices require a robust notion of continuous responsible agency. Some philosophers argue on conceptual grounds that memory cannot be the basis of personal identity because remembering *presupposes* a self who remembers. But others, notably philosophers who see the self as less unified, stable, and integrated than is acknowledged in traditional philosophical theories of personal identity, urge attention to real case studies of, for example, amnesia and dissociative identity, or to better cognitive theories of the selective and constructive nature of autobiographical memory (Schechtman, 1994).

### Memory and Time

What role does memory play in our understanding of time? How is time represented in memory? The philosopher John Campbell (1997) has developed a new picture of the structure of time in autobiographical remembering. For Campbell, the human ability self-consciously to identify personal episodes as having happened at particular past times is bound up with our unique capacities to locate events, and ourselves, in an asymmetric temporal and causal order. Just as our spatial representational skills are not restricted to egocentric models, so our mature temporal orientation involves an objective conception of time as linear. Drawing on empirical work on the representation of time in humans and other animals, Campbell claims that only humans are genuinely oriented with respect to particular times rather than merely to phase or rhythm. Because we can grasp the temporal relations between temporal cycles or phases, we have a conception of the connectedness of time which gives us the concept of the past. This detached or reflective sense of time is what grounds our understanding of the uniqueness of particular actions.

Campbell's audacious analysis of memory and time suggests, among other things, that there is no true episodic memory in nonhuman animals, and thus offers both empirical and conceptual challenges to his opponents. Comparative ethology and cognitive anthropology are as relevant to the evaluation of his account as is the clinical neuropsychology of amnesia. So this case can serve as a first illustration of the possibility that the immediate future of the cognitive philosophy of memory science will be bewilderingly and excitingly interdisciplinary.

### Causation, Representation, and Traces

For me to have a genuine personal episodic memory, my present act of remembering must be causally connected in an appropriate way to the past experience being recollected. Even if it happens to be true that, as a child of four, I got lost in a shopping mall, most of us would deny that I truly personally remember the experience if I had forgotten it, and have only later been told about it by my parents. For this reason, philosophers and psychologists have hypothesized the existence of some kind of 'memory trace' as a continuous physical bridge across the temporal gulf, causally connecting past and present.

When I remember an episode of my personal history, I come into contact with events and experiences which are no longer present. We find it easy to engage in the peculiar sort of 'mental time travel' involved in such autobiographical memory, although we're often aware of significant limits to its reliability. Remembering is an instance of a general, flexible human capacity to think about the absent, so that mental life isn't entirely determined

by the current environment and the immediate needs of the organism.

This is one intuitive route to the reliance on mental representations which lies at the foundation of cognitive science. It's natural to think that the physical trace, which is itself the causal result of past experience, somehow 'represents' that experience, or at least carries sufficient information about the past to allow the organism now to reconstruct that experience or something like it. Since we are often able to remember *without* having any such traces in our current external environment (such as photographs or words written in a diary), many philosophers and scientists have postulated memory traces in the brain. There are stronger and weaker versions of this view.

## COGNITIVE SCIENCE AND THE PHILOSOPHY OF MEMORY

### Localist Models of Memory

Since many historical theorists of memory, from Aristotle to Descartes and on into the twentieth century, explicitly refer to memory traces, it may seem that little progress has been made: 'it is the survival of the memory "trace" concept, some static, permanent, distinct storage form that each experience leaves in the organism, that links together most remarkably the oldest and most modern models' (Colville-Stewart, 1975, p. 402). On this version of the memory trace hypothesis, traces must be independent, 'atomic' items, laid down separately by every experience (or perhaps every part of every experience), and stored at a separate location, until called out again in the reproduction of that experience.

This 'archival' or 'localist' model of information storage has to some extent been physically realized in the design of digital computers since the work of von Neumann in the 1940s. At an abstract level it underlies influential models of human memory in cognitive psychology in which data are clearly separated from processing, or storage system from executive. But it has serious conceptual problems. Followers of the philosopher Ludwig Wittgenstein pursue his point that such static, structural traces are not required by empirical evidence: 'nothing seems more possible to me than that people some day will come to the definite opinion that there is no copy in either the physiological or nervous systems which corresponds to a particular thought, or a particular idea, or memory' (quoted in Stern, 1991, p. 208). Phenomenological and direct realist philosophers, in turn, argue that the view collapses into either incoherence or skepticism. Either, in searching through stored items for the representation of a particular past experience, we must already remember that past experience in checking which is the right representation, in which case the postulation of the trace is redundant; or, if there is no such independent access to the past, then we are trapped in the present behind a veil of memory ideas, and may not really know the past at all (Wilcox and Katz, 1981).

These anti-representationist criticisms are behind many philosophical doubts about the cognitive science of memory, and sometimes about scientific psychology as a whole. But unfortunately such critics have rarely offered positive alternative approaches, beyond the blunt direct realist claim that memory is an immediate, noninferential awareness of past things themselves. And in fact only strongly localist accounts of fixed and static memory traces are vulnerable to these charges. Alternative, more dynamic conceptions of the memory trace may be more empirically plausible, while yet offering genuinely causal accounts of the mechanisms linking past and present.

### Distributed Models of Memory

'Distributed' models of the memory trace build in a degree of plasticity. In connectionist cognitive science, for example, occurrent remembering is the temporary activation of a particular pattern or vector across the units of a neural network. This reconstruction is possible because of the conspiring influences of current input and the history of the network, where this history is sedimented in the particular connection weights between units. Memory traces then are not stored statically between experience and remembering, but are piled together or 'superposed' in the same set of weights (McClelland and Rumelhart, 1986, p. 193):

> We see the traces laid down by the processing of each input as contributing to the composite, superimposed memory representation. Each time a stimulus is processed, it gives rise to a slightly different memory trace – either because the item itself is different or because it occurs in a different context that conditions its representation … the traces are not kept separate. Each trace contributes to the composite, but the characteristics of particular experiences tend nevertheless to be preserved, at least until they are overridden by canceling characteristics of other traces. Also, the traces of one stimulus pattern can coexist with the traces of other stimuli, within the same composite memory trace.

This framework postulates two abstract features: distinct transient patterns of activity, and

composite, enduring, but modifiable dispositional states. It is not tied to current computational models, for these two features can be implemented in different physical systems. But how do such distributed models of memory escape the philosophical criticisms of local traces? Firstly, they offer an account of causal continuity which doesn't rely on the permanent storage of independent items, and which thus seems compatible with Wittgenstein's query 'whether the things stored up may not constantly change their nature' (quoted in Stern, 1991, p. 204). Secondly, they suggest a fallibilist realist response to the skeptical worry: while past experience exerts a robust causal influence on the holistic system of traces, veridical recall is not (and need not be) guaranteed. Finally, they can be integrated into a broader consensus in cognitive psychology about the reconstructive nature of memory, to which we now turn.

## Constructive Remembering

'A variety of conditions exist', notes the psychologist Daniel Schacter, 'in which subjectively compelling memories are grossly inaccurate' (1995, p. 22). Partly in response to the crisis of the early 1990s over recovered memories and false memories, cognitive psychologists have recently developed a striking consensus about the extent and importance of memory distortions and confusions, assuming that understanding mechanisms of distortion will also elucidate the reliability of memory. Research in a number of areas has focused on the context of retrieval, describing the shaping influences of the remembering situation: many memories are created at the moment they are needed, not simply extracted whole from storage. Related work on source monitoring, interference, and suggestibility has developed fairly independently of connectionist computational modeling. This makes the current state of the sciences of memory a key case study in philosophy of cognitive science.

Because memory is studied in many different disciplines, from neurobiology to narrative theory, there is no obvious unity to either the objects of enquiry or the methods employed. In addition to analyzing evidence for particular psychological classifications of the variety of memory systems, seeking for example the best definition of the notion of episodic memory, a further central task for philosophy is the careful evaluation of relations between different levels of explanation. Memory is thus a test case for the possibility and the pitfalls of genuine interdisciplinarity in cognitive science. The cognitive sciences of memory are as yet immature, yet they harness the vast institutional apparatus of normal science.

Those who value theoretical coevolution across disciplines often optimistically underestimate the difficulty of translating terms and postulated mechanisms; while more skeptical voices doubt the likelihood of mutually constraining explanations across neighboring levels or disciplines. Yet there are models of interdisciplinarity in the philosophy of science which allow for contact or even local reductive identification between theories at different levels, without unrealistic reliance on the formal unity of science. Valerie Hardcastle, for example, argues (1996, pp. 105–139) that consensus on the existence of dual implicit and explicit memory systems exemplifies a developing 'complicated and cluttered' interdisciplinary theory which relies actively on the methods and underlying assumptions of a number of different research traditions, in this case including developmental psychology, clinical neuropsychology, and experimental cognitive psychology.

The case of autobiographical memory development, to take another example, suggests that this multidisciplinary reach must extend beyond the level of individual psychology to incorporate social and historical factors. Although children talk about the past almost as soon as they start talking, they may first learn some personal narrative forms to organize their explicit, *public* recounting of past experiences, and then use these principles as scaffolding around which to organize their *internal* representations of past experiences. So the development of autobiographical conventions and patterns of coherence is less an automatic unfolding of internal processes than an internalization of cultural schemes (Nelson and Fivush, 2000). The best explanations of the form and content of specific personal memories may often refer not simply to the past episode itself, but (as in the case of failures of source memory) to multiple causes which span internal and external factors. The converse point also applies. This need to bridge the 'personal' level of explanation, the traditional domain of philosophy of mind, and to focus simultaneously on the subpersonal and the social, has implications for the philosophy of the social sciences of memory, in which explanation may need to refer to flexible internal processes of schematization and reconstruction.

Both cognitive anthropologists and philosophers drawing on dynamical and situated approaches to cognition have suggested the need for a general framework for memory science which can make sense of traces both inside and outside the

individual. This is not to collapse the distinction between external and internal representational formats: for a connectionist in particular, the kind of 'storage' mechanisms employed by the brain are quite different from those of most external linguistic or digital systems. The point rather is to see brain traces and external traces as parts of temporarily integrated larger systems, used by us so as more successfully to exploit and manipulate information in the environment. As Andy Clark puts it, 'our brains make the world smart so that we can be dumb in peace' (1997, p. 180). This perspective can be given a biological twist, as in Merlin Donald's provocative account (1991) of the crucial changes in cognitive architecture which occurred with the historical development of enduring and transportable external memory hardware or 'exograms'; a psychological twist, as in David Rubin's (1995) ambitious theory of memory for oral traditions; or a metaphysical twist, as in perceptual psychologist Michael Leyton's (1992) general theory of memory as the asymmetric traces left by processes on objects. In each case, the understanding of changing cultural and technological systems becomes an integral part of cognitive science, rather than a humanistic curiosity.

## References


Campbell J (1997) The structure of time in autobiographical memory. *European Journal of Philosophy* **5**: 105–118.

Clark A (1997) *Being There: Putting Brain, Body, and World Together Again*. Cambridge, MA: MIT Press.

Colville-Stewart SB (1975) *Physico-Chemical Models of the Memory Storage Process: The Historical Role of Argument from Analogy*. PhD thesis, University of London.

Donald M (1991) *Origins of the Modern Mind: Three Stages in the Evolution of Culture and Cognition*. Cambridge, MA: Harvard University Press.

Hardcastle V (1996) *How to Build a Theory in Cognitive Science*. Albany, NY: State University of New York Press.

Leyton M (1992) *Symmetry, Causality, Mind*. Cambridge, MA: MIT Press.

McClelland JL and Rumelhart DE (1986) A distributed model of human learning and memory. In: McClelland JL and Rumelhart DE (eds) *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, vol. 2, pp. 170–215. Cambridge, MA: MIT Press.

Nelson K and Fivush R (2000) Socialization of memory. In: Tulving E and Craik FIM (eds) *The Oxford Handbook of Memory*, pp. 283–295. Oxford, UK: Oxford University Press.

Rubin DC (1995) *Memory in Oral Traditions: The Cognitive Psychology of Epic, Ballads, and Counting-out Rhymes*. Oxford, UK: Oxford University Press.

Schacter DL (1995) Memory distortion: history and current status. In: Schacter DL (ed.) *Memory Distortion: How Minds, Brains, and Societies Reconstruct the Past*, pp. 1–43. Cambridge, MA: Harvard University Press.

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

Stern DG (1991) Models of memory: Wittgenstein and cognitive science. *Philosophical Psychology* **4**: 203–218.

Wilcox S and Katz S (1981) A direct realist alternative to the traditional conception of memory. *Behaviorism* **9**: 227–239.


## Further Reading


Campbell J (1994) *Past, Space, and Self*. Cambridge, MA: MIT Press.

Casey ES (1987) *Remembering: A Phenomenological Study*. Bloomington, IN: Indiana University Press.

Connerton P (1989) *How Societies Remember*. Cambridge, UK: Cambridge University Press.

Deutscher M (1989) *Remembering 'Remembering'*. In: Heil J (ed.) *Cause, Mind, and Reality*, pp. 53–72. Dordrecht, Netherlands: Kluwer.

Draaisma D (2000) *Metaphors of Memory*. Cambridge, UK: Cambridge University Press.

Engel S (1999) *Context is Everything: The Nature of Memory*. New York, NY: WH Freeman.

Hacking I (1995) *Rewriting the Soul: Multiple Personality and the Sciences of Memory*. Princeton, NJ: Princeton University Press.

Hamilton A (1999) False memory syndrome and the authority of personal memory-claims: a philosophical perspective. *Philosophy, Psychiatry, & Psychology* **5**: 283–297.

Hoerl C (1999) Memory, amnesia, and the past. *Mind and Language* **14**: 227–251.

Hoerl C and McCormack T (eds) (2001) *Time and Memory: Philosophical and Psychological Perspectives*. Oxford, UK: Oxford University Press.

Krell DF (1990) *Of Memory, Reminiscence, and Writing: On the Verge*. Bloomington, IN: Indiana University Press.

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

Rowlands M (1999) *The Body in Mind: Understanding Cognitive Processes*. Cambridge, UK: Cambridge University Press.

Sorabji R (1972) *Aristotle on Memory*. London, UK: Duckworth.

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

Sutton J (2002) Representation, reduction, and interdisciplinarity in the sciences of memory. In: Clapin H, Staines P and Slezak P (eds) *Representation in Mind*. Westport, CT: Greenwood Publishers.

Warnock M (1987) *Memory*. London, UK: Faber.