Causal Theories of Memory

Causal theories of memory aim to give a philosophical account of what it is to remember something. According to such theories, to remember something is to be in a mental state or undergo a mental episode that stands in an appropriate kind of causal connection to one or more of one's previous mental states or episodes. More specifically, causal theories of memory typically have it that the causal connection involves the persistence of a *trace*: the original experience of an event has left a trace on the subject, which is now operative in the subject's remembering the event. In psychology, the idea that remembering involves the activation of memory traces is typically taken as uncontroversial, the main focus of research being on what is known as the problem of localization, i.e. of trying to identify brain structures that realize such traces. In fact, however, attempting to explain *what it is* to remember in terms of the idea of a memory trace turns out far from straightforward. This entry summarizes arguments for and against causal theories of memory and distinguishes two ways of understanding the appeal such theories make to the notion of a trace.

The Basic Intuition and the Problem of Deviant Causal Chains.

Causal theories of memory typically take as their remit the concept of remembering quite generally, but most of the more detailed analyses focus on remembering particular, personally experienced events, or what psychologists call *episodic memory*. As applied to episodic memory, one basic intuition that informs causal theories of memory is that there can be cases in which a subject has in fact experienced a certain past event and can now represent it correctly, without it being true that she remembers the event. For instance, a subject might no longer remember her first kiss, yet nevertheless represent it correctly because she is reading a diary she

kept at the time. The reason why this does not constitute a case of remembering, according to the causal theorist's diagnosis, is that the subject's current ability to represent the event does not stand in the right sort of causal connection to her having experienced it.

Obviously, there remains a need for the causal theorist to elucidate what is meant by 'the right sort of causal connection' in this context, and what, for instance, disqualifies the kind of causal connection that obtains in our example, where the diary forms the causal link between past and present. This is an instance of the problem of *deviant causal chains*, which also affects other causal theories in the philosophy of mind (e.g. of perception and action). We can distinguish between two ways of construing causal theories of memory in terms of the type of responses to this problem that they embody.

Memory Traces as Internal

One type of response to the problem of deviant causal chains, in the case of causal theories of memory, involves adding further constraints to the basic idea behind such theories. One such constraint might be that memory necessarily involves traces internal to the body of the subject, which would rule out the diary entry as a suitable trace.

This response turns on an understanding of causal theories of memory, according to which they imply some substantive assumptions about the nature of memory traces, for instance that they must be realized in the subject's internal physiology (the technology of the day is often invoked to illustrate the idea of a trace in this sense: e.g., wax tablets, phonographic records, or connectionist networks). Yet, the more causal theories introduce such substantive assumptions, the less they seem

able to offer a plausible analysis of our common-sense concept of remembering, which is typically taken to be their aim. It does not seem true that having the concept of remembering requires having beliefs about physiology, any more than having the concept of talking to someone on a telephone requires having beliefs about the processes which make this possible. Thus, this way of construing causal theories of memory lays them vulnerable to the charge of scientism, i.e., of confusing empirical hypotheses about mechanisms underpinning memory with an insight into what it is to remember.

The Very Notion of a Trace

A second type of response to the problem of deviant causal chains, in the case of causal theories of memory, focuses on the notion of a trace itself, and the way it figures in the theory. Going back to our example, there is an intuitive sense in which the diary entry is not a direct trace of the subject's experience of her first kiss; there is further work for the subject to do, in addition to experiencing the kiss, for the diary entry to be produced. Similarly, it is not the diary entry as such, but the subject's reading it, that enables her to represent the kiss. So we can draw a distinction on structural grounds between this case and a case of genuine remembering.

Unlike the first type of response to the problem of deviant causal chains discussed above, this response does not introduce substantive assumptions about the nature of memory traces. Rather, it looks at the ontological categories (such as that of an event, an ability, etc.) that we need to invoke in order to get clear about the nature of remembering, and understands the notion of a trace as one such category. This way of construing a causal theory of memory can perhaps best be understood by contrasting it with a rival view, according to which remembering should be analysed

in terms of the idea of the retention of an *ability* to represent the past, where this is explicitly to be contrasted with the idea of the persistence of a trace. Indeed, on that rival view, it is only in cases in which we *fail* to remember events that we have to rely on traces, as for example when we have to rely on a diary entry to represent our first kiss. One problem with this suggestion is that traces such as diary entries often do play a crucial role in bona fide remembering, by acting as a *prompt* (or *retrieval cue*). Consider again the example of a subject reading about her first kiss in her old diary, but not being able to remember the kiss. Contrast this with an alternative scenario in which the memory of her first kiss in fact comes back to her upon reading the diary. To spell out what constitutes the difference between the two cases, it seems that we need to introduce the idea that, in the second case, something other than the diary entry *grounds* the subject's ability to represent the event. Thus, it is not obvious that there is a viable alternative to causal theories of memory that recruits only the notion of the retention of an ability without presupposing the idea of a memory trace grounding this ability.

Christoph Hoerl

See also: Memory; Autobiographical Memory; Episodic Memory, Computational Perspectives.

Further readings

Malcom, N. (1963). Knowledge and Certainty. Englewood Cliffs, NJ: Prentice Hall.

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

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

Memory and Knowledge

One important strand of psychological research on memory has been concerned with uncovering ways in which human memory can be unreliable, misleading, or even involve complete fabrication. In apparent contrast to this line of research, philosophical discussions of memory typically seek to give it a central, indispensable role in knowledge. This entry will review some of the accounts philosophers have given of the epistemology of memory, before briefly returning to the question as to the potential relevance of empirical research in psychology for such accounts.

What role(s) does memory play in knowledge?

To bring out one way in which memory may be thought to be central to knowledge, consider the following argument attacking the so-called 'myth of the given'. Its target is a position in epistemology known as classical foundationalism, according to which empirical knowledge must ultimately rest on a set of 'basic beliefs' whose epistemic status does not, in turn, depend on that of other beliefs. Traditionally, foundationalists' favourite candidates for such basic beliefs have been beliefs about our own sensations. Suppose you have a certain visual experience on the basis of which you form the belief 'There is a ripe tomato in front of me'. Arguably, the epistemic status of that belief turns on whether you are right to believe that things actually are as they visually appear to you. Yet, the foundationalist would claim that there is another belief you can form in this situation, whose epistemic status does not seem to turn on that of other beliefs, namely the belief that you are, at any rate, having a red sensation.

Implicit in this line of thought is the idea that the mere having of a sensation, by itself, can put you into a position to have a belief about it – and this is what has been criticised as the 'myth of the given'. Your having the belief 'a red sensation is occurring' or even just 'this type of sensation is occurring' seems to require that you can think of your current sensation as being of one type *rather than some other*, which in turn seems to require drawing on memory. If you can't remember (and thus have beliefs about) any other sensations you could have instead, your putative beliefs about your current sensation will be devoid of content – there will be nothing in them that can distinguish that sensation from any other sensation. Thus, it looks as though not even beliefs about our own sensations can serve as 'basic beliefs' in the foundationalist's sense.

It is often said that, without memory, we would know very little, because any knowledge we might have through sensory experience would only last as long as the experience itself. The above example suggests that this statement might not go far enough in acknowledging the epistemic centrality of memory. Rather, if the kind of attack against the 'myth of the given' sketched is along the right lines, memory plays a key role in our very ability to gain knowledge from experience. At the same time, however, there is of course also a sense in which memory, in turn, depends on a capacity for experience (or other capacities for acquiring knowledge). Memory is not itself a faculty for coming to know something; it is dependent on there being other such faculties. As it is sometimes put, memory is not a source of knowledge, or, if it is, it is a *preservative*, rather than a *generative* source.

Memory and justification

Many epistemological theories are focussed primarily on generative sources of knowledge, and, as a consequence, at least some of them have difficulties accounting for the distinctive epistemological significance of memory. For instance, there is an influential tradition in epistemology that is centred of the notion of *justification*, often associated with the thought that justification is a necessary condition for knowledge (as opposed to, say, mere true belief). Following this tradition, it is sometimes assumed that questions about the epistemological significance of memory are best approached by asking in virtue of what beliefs retrieved from memory (henceforth: *memory beliefs*) might count as beliefs the subject is justified in holding. On closer inspection, however, it is not obvious that this latter question best captures the role memory plays in knowledge.

Consider one possible answer to the question as to what justifies memory beliefs: that remembering that *p* is itself a source of justification for the belief that *p*. In effect, the strategy behind this answer is to downplay the epistemological significance of the generative/preservative distinction: memory may be preservative with respect to *content* – i.e. it preserves beliefs acquired by some other means – but it is generative with respect to *justification*. This latter idea is typically spelled out in terms of the thought that there is a specific phenomenology attached to retrieving beliefs from memory (as opposed to, say, just guessing). There is an *experience of recall*, which can serve as a justification for believing that *p*.

One problem for this view is that it is by no means clear how exactly invoking the idea of a distinctive phenomenology of retrieval might help flesh out the idea that remembering is itself a source of justification. It is perhaps tempting to think that such experiences can play a similar role in the justification of memory beliefs as, say,

visual experiences play in the justification of visually-based beliefs. Yet, intuitively, the epistemic role of perceptual experiences has something to do with the fact that there is a sense in which those experiences directly present us with the very things our beliefs are about. In particular, it is the specific perceptual experience I have which makes it rational for me to form the specific beliefs I do. By contrast, the putative epistemic role of memory beliefs, on the view we have been considering, would have to be rather different. In so far as there is a distinctive experience of recall, it seems to be the same experience that accompanies different instances of memory retrieval, no matter what beliefs are being retrieved.

An alternative answer to the question as to what justifies memory beliefs turns on the idea that memory is not just preservative with respect to the content of beliefs, but also with respect to their justification. As normally understood, the suggestion here is not that, in order to be justified in believing that *p* (where the belief that *p* is a memory belief), I need to be able to remember the circumstances under which that belief was acquired, and thus be in a position to rehearse my original justification for acquiring the belief. This would make all but a fraction of our beliefs epistemically problematic. Rather, as it is normally understood, the view at issue here is that beliefs held in memory in fact retain the justification with which they were originally acquired, even if the subject herself is no longer able to remember how she acquired the belief.

This view faces the problem that it is not obvious what exactly the idea of a belief's retaining its justification comes to. This idea seems to presuppose that justification is something akin to a property of beliefs, i.e. *states* (of believing that *p*, or *q*, etc.) that a subject is in over time. Yet, when the notion of justification is explicated in the epistemological literature, it is typically by means of examples in

which a subject acquires a belief for the first time, or holds on to a belief in the presence of countervailing evidence. That is, the notion of justification is attached to aspects of the subject's cognitive *activity*, something the subject does at a time.

As already indicated, perhaps the most basic worry about both of the views outlined above is whether it is right to assume that the epistemic significance of memory is best framed in terms of the notion of justification. As against this assumption, some authors have held that it is actually more intuitive to think of the role that memory plays in knowledge in terms of the idea that memory *frees* the subject from the need to seek justification for certain of her beliefs. Any sort of sustained rational enquiry seems to presuppose that we can normally rely on beliefs we acquired earlier without constantly having to establish their epistemic credentials anew. Thus, it might be thought that there is a sense in which memory has a more fundamental epistemic role to play than can be explained by invoking the idea of memory generating or preserving justification for individual beliefs held in memory. Rather, on this view, the epistemic significance of memory needs to be seen within its wider role of making it possible for us to acquire extended bodies of knowledge.

'False memories' and the epistemology of episodic memory

The above considerations are all concerned (at least primarily) with *factual* or *semantic memory*, i.e. the ability to retain knowledge of facts, concepts, or meanings that we learned about in the past, but not necessarily knowledge about the past itself. Yet, perhaps the first thing to come to mind when the issue of the relation between memory and knowledge is raised is the idea that memory plays a role specifically in our *knowledge about the past*. Philosophical discussion of this idea has centred primarily on the epistemology of *event* or *episodic memory*, i.e. the type of memory

for particular past events that we might express, e.g., by saying "I remember seeing/doing *x*". In particular, theorists have tried to reconcile two intuitions about this type of memory: that it plays a fundamental role in our knowledge of the past, and that it involves the having of memory images.

The idea of a memory image is meant to capture a sense in which recollecting specific events from one's past life involves something akin to re-experiencing them. Yet, it has been argued that the having of a memory image – i.e., a present occurrence – cannot constitute our most fundamental way of knowing about the past. The thought has been that if we did not have a more fundamental way of knowing about the past, not involving imagery, we would never come to connect present memory images with the past. This argument, though, seems to assume that the only role imagery might play in knowledge about the past is by serving as *evidence* on the basis of which we make judgements about the past. It can be avoided if we can make sense of an alternative way of viewing the epistemic role of memory images. Specifically, it has been suggested that, in the case of episodic memory, the subject's ability to call up a memory image is itself the specific *form* her knowledge of the past takes.

It is in this context, in particular, that empirical work on the reliability of memory might be thought to raise challenges for the epistemology of memory. Space prohibits a proper review of the large variety of empirical studies in this area. However, very broadly, a general theme that emerges from much of this research is that subjects' memories about past events are susceptible to interference from information received some time after the relevant events took place. At the extreme, entirely false apparent memories can be 'planted' in subjects by giving them misleading information. This clearly raises a general challenge for epistemologists to provide grounds for thinking that, by and large, we can nevertheless regard memory

as reliable. More specifically, though, the intuition that episodic memory has a distinctive epistemological role to play seems to trade on the idea that there is an essential difference between simply remembering facts *about* a past event and recollecting the event itself, i.e. having some more direct access to the past through having witnessed it. It is this idea, which is arguably a key ingredient of our common-sense understanding of memory, that might be seen to be under threat once it is clear just how much what we seem to remember having experienced can actually be the result of post-event construction.

Christoph Hoerl

See also: Autobiographical Memory; Knowledge by Acquaintance; Memory; Memory, Interference with

Further Readings:

Huemer, M. (1999). The problem of memory knowledge. *Pacific Philosophical Quarterly, 80*, 346–357.

Loftus, E.F. (2005). Planting misinformation in the human mind: A 30-year investigation of the malleability of memory. *Learning & Memory, 12*, 361-366.

Martin, M.G.F. (2001). Episodic recall as retained acquaintance. In C. Hoerl & T. McCormack (Eds.): *Time and memory: Issues in philosophy and psychology*. Oxford: Clarendon Press, pp. 257-284.

Neisser, Ulric. (1981). John Dean's memory: A case study. *Cognition, 9*, 1-22.

Owens, D. (1999). The authority of memory. *European Journal of Philosophy, 7*, 312-29.

Soteriou, M. (2008). The epistemological role of episodic recollection.

Philosophy and Phenomenological Research, 77, 472-492