

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.