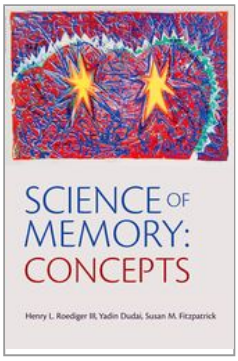

Science of Memory: Concepts
Henry L. Roediger, Yadin Dudai, and Susan M. Fitzpatrick

Print publication date: 2007
Print ISBN-13: 9780195310443
Published to Oxford Scholarship Online: May 2009
DOI: 10.1093/acprof:oso/9780195310443.001.0001

Forgetting

Michael Davis (Contributor Webpage)
Elizabeth F. Loftus
David C. Rubin
John T. Wixted

DOI:10.1093/acprof:oso/9780195310443.003.0015

Abstract and Keywords

This part presents four chapters on the concept of forgetting. The first chapter analyzes the term “forgetting”. The second discusses the impact of misinformation on the ability to remember previous event details. The third considers whether forgetting is a useful concept in the science of memory. It argues that it is not an especially useful in terms of what it denotes, but that what it connotes needs to be kept. The fourth presents a synthesis of the chapters in this part.

Keywords: memory, remembering, misinformation, recall, previous event details

Forgetting is the loss of learned information over time. In general, the forgetting curve, or the function relating performance on a task from time since learning it, is negatively accelerated, with greatest forgetting occurring soon after learning and more gradual loss later. In most modern treatments of the subject, there is no assumption that forgetting necessarily reflects complete loss of stored information; rather, the problem may be loss of access to stored information because of a shift in context between learning and testing, or other reasons. Another definition of forgetting is the inability to retrieve information at a certain point in time that could be retrieved at a previous time. Some studies have shown dramatic recoveries of seemingly forgotten information when appropriate cues are provided to aid retrieval. Of course, the possibility that

forgetting is at least partly due to decay or other types of change of the memory trace is quite plausible and is in fact assumed by most neurobiological theories of forgetting. Mechanisms discussed in theories of forgetting include changes in the nature of the memory trace, interference from earlier and later experiences, change in context between learning and testing, and insufficient or inappropriate retrieval cues.

H.L.R. **(p.316)**

(p.317) 54 Forgetting: Once again, it's all about representations

The term 'forgetting' is often used in different ways, which makes it an inherently difficult concept to study in a rigorous way. One use of the term is to describe the theoretical possibility that refers to a total erasure of the original memory that cannot be recalled, no matter what techniques are used to aid recall. I will call this the 'strong form of forgetting'. The other use of the term is to describe a failure of retrieval which, in essence, refers to all of the other instances where a memory may still exist, but, for whatever reason, cannot be retrieved.

Regarding the strong form of forgetting, the problem is that the only way to determine definitely if a memory is forgotten because of a total erasure of the original memory substrate is to examine an extremely simple neural circuit which can reveal all of the cellular and molecular events that occur when a memory is formed and then to show that all of these events have gone back to their original state at the time when the memory is not retrieved. I say *all of the* cellular and molecular events have to return to their original state because it is conceivable that any residue of change might be enough of a substrate to allow some sort of retrieval or saving, given clever enough experiments designed to bring out the original memory. Only when all the cellular and molecular events that occur when a memory is formed return to their original state would I say this would be evidence for true forgetting.

This strong version of the definition of forgetting could equally apply to short-term versus long-term memory. For example, imagine a very simple organism where short-term memory, defined as a change in synaptic strength, was mediated by phosphorylation of a single protein at a single site, and no other changes were observed. If this site was later dephosphorylated so that the protein returned to exactly its state prior to formation of short-term memory, I would call this an example of the strong form of forgetting. If a long-term memory in this simple organism resulted in the formation of a new synapse, and nothing more, then the loss of this extra synapse would return **(p.318)** the system back in its entirety to that measured prior to memory formation and I would call this the strong form of forgetting. The obvious problem with this strong definition is that it is presently impossible to make these kinds of measurements on any organism, and so we still cannot be sure that forgetting

actually occurs, given these criteria. Moreover, because in a slightly more complex organism we may never be able to make these measurements, I assert that the concept of this strong form of forgetting is not useful scientifically, because it probably can never be proven.

Another way to look at the strong form of forgetting would be with respect to dementias associated with cell loss. Basically, if all the cells in a complex network that stored a particular memory finally die, then I would consider this to be forgetting, because there would no longer be any substrate to allow for subsequent retrieval. I watched my mother gradually lose her memory over the last 10 years of her life. I used to test her when I saw her and asked her where her mother was born. Even though she no longer knew who I was, as best as I could tell, she would confidently say: Dearborn, Michigan. By the time she was 90, she no longer could remember even this very dear memory and her MRI showed there were no cortical cells left. So, I would say this would be an example of the strong form of forgetting. However, other than studying why cells die and how to prevent them from dying, this definition of forgetting becomes sort of trivial and self-evident, so again not very useful for the study of memory.

This leaves us with the second use of the term forgetting that deals with ‘retrieval failure’, a term I favor because it instantly invites the possibility of multiple mechanisms to account for an empirical phenomenon, namely one where a memory is absent, incomplete, inaccurate or even false. Moreover, this is more or less how most people regard forgetting and jibes with their personal experience (I cannot come up with his name right now). Numerous mechanisms have been used to explain such instances of retrieval failure, such as repression, interference, extinction, context, response competition, unlearning, pattern completion failure, as well as the strong form of forgetting, namely a full erasure of the original memory. So, this is a better term because it encompasses a larger number of possible, and researchable, mechanisms and does not have the connotation of an erasure of the original memory, although it does not exclude this as a possible mechanism. Also, it is clear there is a rich and very important literature looking at all of these mechanisms with respect to memory retrieval. My position is that it would be much better to view this literature in terms of factors involved in memory retrieval, or perhaps remembering, which are not necessarily identical concepts, rather than mechanisms of forgetting. However, implicit in this position is the idea that memories may never be **(p.319)** ‘forgotten’ in the strong sense of the term, but only inaccessible, as a result of one or more of these mechanisms. However, this also seems not to be very satisfactory because it would seem useful for a brain not to remember everything it ever learned, given that this could lead to more retrieval failure based on interference mechanisms.

For example, if I travel a lot and stay in lots of hotels, I may at some time ‘forget’ the room number of the hotel where I am staying on a particular night. This would be forgetting during temporary long-term working memory and would probably be an example of retrieval failure because it is likely I could retrieve my room number with priming, recognition tests, etc. (i.e. if I was taken to the floor of my room I might well be able to figure out where my room was, based on the proximity of the elevator or other cues). However, if one makes the assumption, and it is only that, that the room number of every hotel I have ever stayed in during my life was never put into long-term memory, but only in some kind of temporary memory buffer accessible more or less during the time of my stay in the hotel and not much beyond that, then I think it is not useful to say that ‘I forgot the room number of the hotel I stayed in 30 years ago’, using the strong definition of forgetting. Furthermore, I do not think it is useful to say that ‘I fail to retrieve the room number of the hotel I stayed in 30 years ago’ because this implies the long-term memory was laid down in the first place, which I also think is unlikely, i.e. I would not like to say I have forgotten those room numbers or even that I fail to retrieve them. Instead, I would say I do not retrieve them now because those temporary working memories purposely were never consolidated into long-term memory.

So, where does this leave us? To those of us interested in memory and forgetting, I think it is critical first to ask: what are the possible sources of forgetting in the model of learning we are looking at. Could it be explained by a total erasure of the original memory, a partial erasure that would be enough to prevent retrieval permanently because it would have degraded the original pattern of the memory to such an extent that it can never be retrieved, a temporary retrieval failure that could be retrieved at some later time or with some other cues, or because the memory was never put into some form of long-term storage. Perhaps by always asking these questions we might be able to design better experiments to try to answer these somewhat different questions. **(p.320)**

(p.321) 55 Forgetting: The fate of once learned, but “forgotten” material

Sometimes an experience is stored in memory, and later becomes difficult, if not virtually impossible, to remember. We can only speculate about whether it is ever completely gone from the memory into which it was once stored.

A different situation arises, however, when an experience is stored in a person’s memory, and later the person is exposed to some erroneous and contradictory fact. Assume that it too is stored in memory and that the post-event contradictory detail is all that the individual can now remember no matter how hard he or she tries. So, for example, a person sees a car go through a red light but later hears another witness call the light ‘green’. Assume our witness embraces the misinformation and is now confident that he or she saw a green

light. A fundamental question, then is, ‘What happened to the original memory for the red light?’ (Loftus and Loftus 1980).

A major issue that has been debated is whether misinformation actually impairs a person’s ability to remember event details. Put another way, are memory traces actually altered by post-event misinformation? There are several ways in which misinformation could impair the ability to remember event details. First, misinformation could cause ‘trace impairment’, i.e. it could update or alter the previously formed memory. New information could combine with earlier traces actually to change the representation. A second way in which misinformation could impair event memory is through ‘retrieval impairment’, i.e. misinformation may not alter the original memory trace but may simply make it less accessible. (Morton *et al*. 1985). Impairment of some sort is implied by either the trace impairment or retrieval impairment mechanisms. Trace impairment implies real forgetting, however.

Some theorists have rejected the notion that misinformation impairs the ability to remember event details (McCloskey and Zaragoza 1985). They claim there is no impairment of either type in this type of situation, either trace **(p.322)** or retrieval. Michael McCloskey and Maria Zaragoza disagreed with the idea that the misinformation effect was due to recoding processes or updating of previously stored memories, or that the misinformation effect arose because the older memory was rendered less accessible through a mechanism of inhibition or suppression. McCloskey and Zaragoza argued instead that the misinformation does not affect memory at all. Misinformation merely influences the reports of subjects who never encoded (or do not recall) the original event. Instead of guessing at the time of test, these subjects would be lured into ‘guessing’ the misinformation item. Misinformation effects could also be obtained if subjects remember both sources of information but select the misleading information because they conclude it must be correct. While these investigators did not observe a misinformation effect with what they called the ‘modified test’, subsequent investigators would (for a review, see Ayres and Reder 1998).

Several lines of experimental evidence were offered to support the notion that misinformation occasionally does impair the ability to remember original details. First, there are studies using tests that do not permit the misinformation option. Say a subject originally saw a stop sign but it was later referred to as a yield sign. Suppose we now give the subject a test that does not permit the selection of the yield sign (e.g. choose between a stop sign and a no parking sign). If the misinformation impaired memory for the stop sign, then the misinformed subjects would be less able to recover the stop sign (assuming memory impairment had occurred). If there is no memory impairment due to misinformation, then misled subjects would be expected to be as accurate as control subjects on a test of this type. Although some studies do show equal performance, there are several published demonstrations of deficits in

performance with this restrictive type of test. One study presented pre-school children with stories followed by misinformation, and found impairment (Ceci *et al.* 1987). Another study presented adult subjects with visual scenes (e.g. nature scenes including ponds, flowers and mountains), and then provided similar visual scenes as post-event information. Subjects who received misinformation were less able to discriminate the original scene from novel distractors (Chandler 1991).

A second line of work supportive of a memory impairment interpretation involves the use of a yes/no test (Belli 1989). Robert Belli showed subjects a simulated crime, and then fed them some misinformation on a post-event narrative that they read under a pretense. Finally subjects were presented with a series of statements, each dealing with a critical event item. Subjects said 'yes' if they saw the item in the slides, and 'no' otherwise. Compared with memories for control items, there was a large reduction in accurate memory for the **(p. 323)** items about which subjects had received misinformation. The large reduction was not offset by the small improvement in memory (e.g. lower false alarm rate) for completely novel items.

Ayres and Reder (1998) reviewed the various theoretical perspectives that have been used to account for misinformation and its influence on original memories. They concluded that an activation-based semantic network model of memory was most useful in terms of its ability to account for the various and sometimes conflicting results in the literature; and they proposed it as a useful guide for planning for future research in the area.

Another way to think about the issue of whether misinformation impairs a person's ability to remember a previous event is to focus not so much on whether the earlier trace has been altered or degraded but on whether the retrieval cue becomes less effective. Wixted (2005) has referred to this problem as 'cue overload'. He argues that forgetting can involve trade degradation but also that cue overload effects are very real, as has been shown in numerous laboratory studies (e.g. studies involving proactive interference or list length effects, to name a few). Misinformation could be impairing memory performance by effectively increasing cue overload. Since cue overload is such a well-established principle, it seems likely that decrements in memory performance would occur by impairing access to the original information.

The debate over the fate of original memories may not have settled the issue of whether original memories are ever truly destroyed, as this was impossible to prove definitively. However, it did heighten appreciation for the different ways by which people come to report a misinformation item as their memory (Loftus and Hoffman 1989). Sometimes this appears to occur because they have no original memory, either because it was never storied or because it has faded away. Sometimes this occurs because of a conscious deliberation (e.g. I thought I

saw a green light, but the other witness says red, so I'll say red.). Sometimes it appears as if the original event memories have been impaired in the process of contemplating misinformation.

For the sake of historical perspective, it is worth noting that much of the debate about the fate of previously stored memory traces after exposure to misinformation is reminiscent of the debates that occurred a half century ago under the rubric of 'interference theory' (see Roediger 1996). Back then, many researchers were less interested in examining distortions that permeated memory reports than in whether interfering material produced unlearning of prior material or whether it merely added to response competition that resulted in the appearance of 'forgetting'. By the early 1970s, interference theory, once wildly exciting to memory scientists, was becoming less important in the literature, when the first of the hundreds of misinformation studies **(p.324)** began to be published. The new misinformation studies were similar in some ways to the earlier 'interference' studies in that both involved exposure to later material that reduced memory for earlier material. Both paradigms led to extensive discussion of the fate of the earlier material after exposure to later material. However, the misinformation studies often involved situations where new details were constructed in memory, as a result of misinformation, that were never explicitly mentioned.

The study of misinformation effects has had a 30 year history (Loftus 2005), and continues to be studied today. However, research specifically on the issue of the fate of original memories wound down in the 1990s, perhaps because of the enormous challenge of ever being able to prove that the memory trace was truly altered, distorted or gone. Alternatively, perhaps it was because researchers found it of sufficient interest that the memory reports of subjects could be so readily altered that they focused more on the behavior itself and its implications. At this stage, the whole topic of the fate of previously stored memory traces is still, to many, a fascinating and unresolved issue that is ripe for further investigation.

(p.325) 56 Forgetting: Its role in the science of memory

Forgetting is a key concept in the study of memory; it is what concerns people most. Remembering goes unnoticed, only forgetting draws attention. Thus researchers need to maintain forgetting as part of their vocabulary and use it to communicate their work to the public. Forgetting provides a window on memory processes, on why some experiences are lost more easily than others. However, my question is not whether forgetting is a useful concept in general, but whether forgetting is a useful concept in the science of memory. I will argue that it is not an especially useful in terms of what it denotes, but that what it connotes needs to be kept.

There are two ways to consider the concept forgetting. The first is in relation to other concepts; the second is in what we know about forgetting in particular. According to the framework of this book, concepts are mental representations that are always linked to other concepts, all of which are ultimately expressed in language. The key concepts of this book to which forgetting is most closely linked are remembering, retrieval, encoding, learning, memory, consolidation and inhibition. I begin by describing these relationships.

In colloquial terms, forgetting is the opposite of remembering, though in the science of memory remembering may have a more specialized meaning. What one cannot remember, one has forgotten. Both are processes that are measured at a particular time in a particular context. Remembering or forgetting here and now does not guarantee a similar behavior in other contexts, as noted by Tulving and Pearlstone (1966). Successful retrieval is what happens when you do not show forgetting. Failing to retrieve is forgetting, which may be remedied by additional retrieval attempts or a change in the retrieval context. Therefore, forgetting, remembering, and retrieval are conceptual variants of each other in that forgetting can be defined as not remembering, or failed retrieval.

A definitional prerequisite for forgetting is that one has encoded or learned something and so could later have a memory for it. Thus, the concepts of **(p. 326)** encoding, learning and memory are central to the concept of forgetting. Encoding must take place for there to be a memory. Learning is a process that results in the product of ‘having’ a memory. Under this view, memory is reification; process turned into product. Thus memories share properties with most other products, or things, including that they tend to have a single location. One can focus on either the process or the product, but not both at the same time. Researchers who stress one tend to have theoretical difficulties with those who stress the other. The product view is clear at the level of a natural language metaphor; it is easy to understand William James’ attic in which things are placed and later retrieved. In contrast, some find the process view difficult to understand even when presented by good writers who have used it to make widely recognized empirical advances (e.g. Gibson 1966; Skinner 1974; Watkins 1990). When I was in graduate school, learning was something studied by Skinnerians who did not like putting things into the head, or perhaps they just felt that bird brains had no room to store memories given the evolutionary requirement of reducing weight for flight. Memory was something studied by cognitive psychologists who liked filling the expansive human mind with data structures of the kind computer science was inventing at the time (Rubin 1988). Many of these cognitive psychologists were reformed ‘verbal learners’ who traced their intellectual heritage back to Ebbinghaus’s (1885/1964) book titled, *Memory*, rather than Bartlett’s (1932) process oriented book titled, *Remembering*.

I prefer process, in part because it emphasizes the interaction of the organism with the environment, whereas structure emphasizes the memory as a thing located inside the organism. Either view is fine for the relatively constant environment of the laboratory, but process seems better suited for the richer and more varied environment of the world where the match between the context at encoding and context at retrieval (i.e. encoding specificity; Tulving and Thompson 1973) can vary more and where the environment can play a greater role in what is forgotten. For instance, if a building in which you spent time is destroyed or if a friend dies, then there are memories that you could have remembered by visiting the building or interacting with the friend that will now be forgotten because the context and cues for recall are lost. As Proust notes, 'The past is hidden somewhere outside the realm, beyond the reach of intellect, in some material object (in the sensation which that material object will give us) … As for that object, it depends on chance whether we come upon it or not before we ourselves must die' (1913/2000, p. 55).

In any case, you can show you have a memory only by remembering something that you learned. Operationally, the demonstration of encoding, learning, memory, remembering, retrieval and not-forgetting are the same: an organism **(p.327)** did not know something at time 1, was exposed to it at time 2, and displayed that it knew it by some test at time 3. Variations at time 2 would be interpreted in terms of encoding or learning; variations at time 3 in terms of remembering, retrieval or forgetting. However, in all cases, behavior at time 3 is the only evidence that there was encoding, learning, memory remembering, retrieval and forgetting. If there was another exposure to and test at time 4, a learning curve could be plotted or the effects of spaced practice studied. Alternatively, if at time 4 there was only another test, a retention function could be plotted. Thus, in terms of what they really refer to instead of what topics one studies under each, demonstrations of encoding, learning, memory, remembering, retrieval and not-forgetting are conceptual variants of each other.

In contrast, consolidation and inhibition are specific claims about processes that differ from and contribute to our understanding of forgetting. Once a memory becomes consolidated, forgetting slows and is less likely to involve individual aspects of the memory which are more strongly bound into a single unit. Questions about the extent to which forgetting is due to active inhibition of material that would otherwise be known are of great theoretical and practical importance.

I now turn to what is usually considered under the concept of forgetting, i.e. the issues that make up the core of what researchers studying forgetting actually study. One key issue is the time course of forgetting, a topic often referred to as a retention function (Rubin and Wenzel 1996) because the quantity measured is the amount retained, not the amount forgotten. A second issue is what, if anything, is ever completely forgotten or how much can be shown to remain

under the right circumstances and test measures (Chapter 54, this volume). A subissue is whether information is overwritten by new information or remains to be later revealed (Chapter 55, this volume). A third issue is how best to account theoretically for the nature of forgetting, a question which led to (1) the development of one of the most detailed and quickly forgotten bodies of literature on human memory—interference theory, though the issues it dealt with remain in the literature on false memories; (2) theories based on the nature of errors that occur when what is forgotten is replaced by more schematic material (Bartlett 1932); and (3) theories of what motivates forgetting and if motivated forgetting can be viewed as a form of active inhibition (Freud 1901/1960). For all these issues, forgetting tends to be most appropriately set in a trace theory context in which known accurate information is either remembered or forgotten, as opposed to a more constructive theory in which the process of remembering produces a memory that is often good enough, even if not accurate.

(p.328) The general view that I find from the study of forgetting is that forgetting occurs to the extent that what is to be recalled cannot be uniquely determined by the effects of current cuing (Watkins 1990; Rubin 1995); disambiguating items is a key aid to lessen forgetting. One can never tell if anything is ever completely forgotten, only that it is forgotten with the effectiveness of the cues at hand, and this effectiveness can change with time and the state of the individual.

Trying to understand forgetting has led to valuable insights, but is the concept of forgetting needed as a key concept in the science of memory? Studying forgetting is one approach to studying memory. At the behavioral level, it highlights the processes of retrieval, consolidation and inhibition, and the interaction of the organism with the environment. At a neural level, it is a means of examining changes in the organism that result in cues to no longer reactivating what were once the neural traces of a memory. Given that, perhaps forgetting deserves a role in the science of memory. However, we do not need all the concepts of forgetting, remembering and retrieval as well as learning and memory. Others may choose to keep forgetting, but I will continue to emphasize remembering. Remembering emphasizes process versus structure, what is, not what is lacking, and the role of the environment as well as the organism. Moreover, it allows memory to be distributed among many systems of the mind and brain in a highly constructive, but often very faithful, fashion (Rubin 1995, 2005).

Acknowledgments

I wish to thank Dorthe Berntsen, Jennifer LeBreque, Heather Rice, Jennifer Talarico and the group at the Science of Memory meeting for their comments, and NIA grant AG023123 for the time to think.

(p.329) 57 *Integrative comments* Forgetting: It’s not just the opposite of remembering

‘Philosophy’, wrote Wittgenstein (1953), ‘is a battle against the bewitchment of our intelligence by means of language’ (p. 109). Psychology is involved in that battle as well, and most would probably agree that the language of *forgetting* would have outlived its usefulness if it tended to bewitch the intelligence more than it helped us to think about and understand how our memories change over time. One obvious problem with the word *forgetting* is that is has more than one meaning (as noted by Davis in Chapter 54). To some, it simply means being unable to retrieve sought-after information from memory (e.g. ‘I forgot your name’), regardless of why that might be. According to this way of thinking, the word *forgetting* applies if I was once in a position to have encoded the information (e.g. if I was introduced to you at a party) but I am now unable to retrieve it. The information might have been momentarily held in working memory, but it may never have been stored in long-term memory. Alternatively, the trace might have been encoded into long-term memory, but it may now have weakened to the point where it is no longer accessible. Still another possibility is that the available retrieval cues are deficient in some way (e.g. an acquaintance was wearing casual clothes when the introduction took place but is now wearing a uniform). In all of these cases, the word *forgetting* is commonly used, especially by laymen. To the layman, none of these distinctions matter because, in common parlance, forgetting is tantamount to retrieval failure (as noted by Rubin in Chapter 56).

To the psychologists who have long studied the issue of forgetting, however, these distinctions are (or should be) critical. A patient who suffers from bilateral damage of the medial temporal lobes, for example, is unlikely to remember your name a day after meeting you. While that might seem to indicate rapid forgetting (and would be described as such by many), we now know that it probably reflects the failure to lay down a long-term memory trace in the **(p. 330)** first place (Squire 1992). As the term has long been used in experimental psychology, the failure to encode does not count as an instance of forgetting. Similarly, it is now well known that memory is cue dependent, with the most effective retrieval cue being that which was encoded along with the to-be-remembered information (Tulving and Thomson 1973). Other cues will be less effective, but the mere fact that a less effective cue fails to elicit retrieval does not imply that forgetting has taken place, at least not as many psychologists have used that term for more than a century. The phrase ‘cue-dependent forgetting’ (Tulving 1974), which could be applied whenever a cue fails to access an available trace, is probably better termed ‘cue-dependent retrieval failure’ to distinguish it from what I would regard as a more appropriate definition of forgetting. Although Davis (Chapter 54, this volume) seems comfortable subsuming these disparate cases under the concept of forgetting, my own preference would be to establish a more exact definition.

A more rigorous definition of forgetting

When considering the concept of forgetting, an important distinction to keep in mind is whether the memory trace in question is thought to reside in working memory or in long-term memory. The term *forgetting* can be properly applied to each, but it seems important to be clear about which store one is referring to (and to avoid unintentionally intermingling the two). For the most part, I will concern myself here with forgetting from long-term, declarative memory, which refers to encoded traces that are outside the focus of attention (i.e. outside of the window of working memory) and that are reanimated by retrieval cues (as opposed to being refreshed by rehearsal). Note that neuroscientists often partition what I call long-term memory into short-term and long-term components, depending on whether or not the consolidation process has triggered the protein synthesis needed to establish an especially durable trace. However, that distinction is not critical here. For my purposes, long-term memory refers to memories that are not in the focus of attention but that can be revived into conscious awareness by retrieval cues.

As psychologists have generally used the term, *forgetting* from long-term memory refers to the inability to retrieve information that was once consciously accessible under the cuing conditions that are again in effect now. More formally, a particular item *i* might have been learned in association with a particular retrieval cue, q_i. At time *t1*, we find that the probability of retrieving item *i* from long-term memory given q_i—denoted $p(R_{i,t1}|q_i)$—is high. Sometime later, at *t2*, we find that the probability of retrieving item *i* given q_i—$p(R_{i,t2}|q_i)$—is low, i.e. forgetting from long-term, declarative memory is said to occur when $p(R_{i,t1}|q_i) > p(R_{i,t2}|q_i)$. Usually, these probabilities are estimated **(p.331)** by testing memory over a set of items, though I will illustrate the point with reference to the forgetting of a particular item. As a concrete example, imagine having dinner at a restaurant called *Maria's Cocina* on Sunday. On Monday, the word *Maria* might occasion the successful retrieval of *Cocina*, but that same cue might fail to do so a month later. That would qualify as an instance of forgetting because it satisfies the criteria set forth above, i.e. we know that the to-be-remembered item (*Cocina*) was encoded into long-term memory because the retrieval cue (*Maria*) effectively retrieved that information 24 h after encoding (well beyond the window of working memory). Later, that same retrieval cue failed to retrieve the sought-after information. Under conditions like that, one would say that forgetting has occurred, at least according to the definition being proposed here.

Note that this definition of forgetting was implicit from the earliest investigations of memory. The famous savings function reported by Ebbinghaus (1885/1913), for example, illustrates the loss of memory that occurs over time even though the cuing conditions in effect during testing were held constant and even though the to-be-remembered information was undoubtedly encoded into long-term memory. Tulving (1974) appears to have had the same definition in

mind when he said “Forgetting, as defined here, is the inability to recall something now that could be recalled on an earlier occasion” (p. 74). I am defining forgetting in the same way, except that I would add two caveats: (1) that the instance of recall that occurred on an earlier occasion involved retrieval from a long-term memory store (not working memory); and (2) that the retrieval cue is the same on both occasions.

A key point is that the study of forgetting should be specific to one memory system. The definition that I have proposed here is the one that applies to forgetting from long-term, declarative memory. As an example, Wixted and Ebbesen (1991) studied the forgetting of words over a relatively short period of time (40 s), but they construed their experiment as a study of forgetting from long-term memory because an extremely demanding distractor task was in effect throughout in order to prevent any involvement of working memory. Unless rehearsed, a trace in working memory is usually thought to survive no more than about 2 s (Baddeley 1997).

One could, of course, study the rapid forgetting that is observed for traces held in working memory, in which case one would be studying forgetting from that memory system (not from declarative memory). Similarly, one could study the loss of a learned skill over time, in which case one would be studying forgetting from procedural memory. The essential point is that regardless of which memory system is involved, the word “forgetting” applies if performance at *t1* and *t2* involved retrieval from the same system and under **(p.332)** the same conditions. In most studies of forgetting, the system in question is long-term, declarative memory, and for that system forgetting is said to exist when $p(R_{i,t1}|q_i) > p(R_{i,t2}|q_i)$.

Philosophical conundrums to avoid

The concept of forgetting is neutral about *why* $p(R_{i,t1}|q_i) > p(R_{i,t2}|q_i)$. Answering the question of why forgetting occurs once fascinated experimental psychologists, but interest in the question has greatly subsided in recent years. However, the loss of interest seems to be more a function of the field’s disillusionment with a particular theory of forgetting (namely, interference theory) than with any inherent problem with the way in which forgetting is conceptualized. The concept of forgetting did not bewitch the intelligence, but an increasingly baroque interference theory of forgetting may have (Tulving and Madigan 1970).

In searching for the answer to the question of why forgetting occurs, it seems best to avoid spending too much time on issues that are inherently untestable and which quickly devolve into the realm of philosophy. For example, the definition of forgetting presented above includes the assumption that the retrieval cue is held constant over time (but information becomes less accessible anyway). One could worry that the retrieval cue used at *t1* is not *exactly* the

same as the retrieval cue used at *t2* despite a conscientious effort to keep them the same. As the Greek philosopher Heraclitus once said, you cannot step into the same river twice. Indeed, whenever $p(R_{i,t1}|q_i) > p(R_{i,t2}|q_i)$, one can appeal to this idea to explain forgetting. However, doing so resurrects a long abandoned notion, namely that the mere passage of time causes forgetting. In its new incarnation, the theory would be that no two retrieval cues can ever be the same if they are separated in time because the passage of time itself somehow changes them. This account is not unlike the decay theory of forgetting that McGeoch (1932) once famously argued against. According to the theory he challenged, forgetting was explained by the idea that the memory trace decays with the mere passage of time. However, McGeoch (1932) argued that time itself does not cause the trace to decay any more than it causes metal to rust. Specific intervening activities and events, he argued, must instead be responsible. The same argument could be applied to the idea that retrieval cues change with the passage of time. The mere assertion that retrieval cues are changed with time does not explain forgetting. One would need to show exactly how they are changed as time passes and how that affects retrieval.

Another philosophical issue that may be best avoided by experimental psychology concerns the final status of the memory trace, a point that Michael Davis also addresses in some detail. As he points out, an extreme view of **(p. 333)** forgetting would hold that true forgetting involves the complete eradication of the memory trace, either because it faded away altogether or because it was entirely overwritten by a new memory trace. However, it would be difficult to establish the complete absence of a trace because it is always possible that an as yet untried retrieval cue would show that some remnant of the trace is still available. Thus, focusing too much time on this definition of forgetting does not seem to be a fruitful course of action (on that, Michael Davis and I agree).

The difference between remembering and forgetting

Remembering, as that word is commonly used, is said to occur when a cue occasions the retrieval of a prior experience that was encoded into long-term memory, i.e. remembering refers to retrieval success. In everyday language, forgetting is the opposite of remembering because, as indicated above, the term *forgetting* is used synonymously with retrieval failure. However, the more technical definition of forgetting that has been in effect since Ebbinghaus first set out to study forgetting refers not to retrieval failure *per se* but to retrieval failure that occurs despite the fact that the retrieval cue being used now once effectively retrieved the information from long-term memory. Thus, as typically used in psychology, forgetting is *not* the opposite of remembering. On this point, I part company with Michael Davis and David Rubin, both of whom see forgetting as equivalent to retrieval failure (in which case forgetting and remembering are tightly connected, reciprocal concepts, as is true in everyday language). Imagine, for example, an experiment in which one group studies a list

of 20 words presented at a rate of one word per second. A second group studies a similar list presented at the same rate while also being required to complete a concurrent task. The next day, we might find that the first group recalls more words than the second, i.e. we might find that the second group experienced more retrieval failure than the first. Did that group therefore exhibit more forgetting? Not necessarily, because it is likely that they failed to encode many of the words in the first place due to the concurrent task at encoding. An earlier test of memory (e.g. 10 min after learning) might have shown that to be true and might have shown that the rate of forgetting in the two groups was the same despite a considerable difference in the amount of retrieval failure.

These considerations call to mind a common referential practice in the neuroimaging literature that strikes me as being too casual and imprecise. Often, neural activity is measured by functional magnetic resonance imaging (fMRI) during a recognition task. Typically, neural activity associated with hits (target items that are correctly identified as having appeared on the list) is compared with neural activity associated with misses (target items that are **(p. 334)** incorrectly identified as being new). The hits are regarded as the remembered items, whereas items that are missed are often referred to as 'forgotten' items (e.g. Davachi *et al.* 2003). What is unclear, though, is whether these items were ever encoded in the first place. Moreover, on a recognition test, whether an item is a hit or a miss depends in large part on where the subject opted to place his or her decision criterion. Had the subject decided to respond in a somewhat more liberal fashion, the hit rate would have been higher (which means that the miss rate would have been lower), and what was labeled as a forgotten item would have instead been labeled as a remembered item. However, this transformation would have taken place despite the fact that the status of the memory trace and the status of the retrieval cue remained unchanged. All that happened was that the subject decided to respond in a more liberal fashion. The subject's *d'* score (the measure of memory strength) would remain unchanged, and it seems odd to refer to changes in remembering and forgetting taking place despite a constant memory strength. In fact, the terms should not be used in this way. Instead, the field seems best served by reserving the term *forgetting* for items that were once retrievable from long-term memory but no longer are despite using the same retrieval cue in both cases. Applying the term forgetting to misses does not pass this test. Of course, contrasting neural activity for hits versus misses seems appropriate enough, but it is probably better thought of as a comparison between strong and weak memories (not between remembered and forgotten items).

Why do we forget?

With the definition of forgetting from long-term memory suitably framed as $p(R_{i,t1}|q_i) > p(R_{i,t2}|q_i)$, the question then centers on why forgetting happens. Traditionally, the explanations have been that something happens to the retrieval cue (something other than the mere passage of time, that is) or

something happens to the memory trace. Interference theory, which has dominated thinking in experimental psychology for decades, essentially holds that forgetting occurs because, during the time between *t1* and *t2*, the retrieval cue becomes associated with additional items, and the more encoded items a retrieval cue subsumes, the less effective it is at retrieving any one of them. This is the 'cue overload' theory of forgetting (Watkins and Watkins 1975). A great deal of convincing experimental evidence demonstrates that overloading a retrieval cue between *t1* and *t2* will, indeed, cause forgetting to happen. However, as noted by Wixted (2004*a*,*b*, 2005), the evidence that this mechanism accounts for everyday forgetting (i.e. that retrieval cues typically become overloaded with the passage of time) is quite limited.

(p.335) The other idea, that forgetting occurs because the memory trace becomes degraded in some way between *t1* and *t2*, is standard thinking in the field of neuroscience. As argued by Wixted (2004*a*,*b*), the evidence that the formation of new memories serves to degrade recently formed memories that have yet to consolidate is strong. Still, this is an open question that warrants the attention of the field (though, at present, the field's attention mostly lies elsewhere).

As Loftus points out in Chapter 55, both of these accounts remain viable contenders to explain why the introduction of misinformation between *t1* and *t2* makes the originally encoded information less accessible than it otherwise would have been. Loftus notes that the field has moved away from exploring these fascinating theoretical issues in the area of misinformation, and Wixted (2004*b*) noted that the same is true of the entire area of forgetting. It is unfortunate that, with a few notable exceptions (e.g. Pavlik and Anderson 2005), the field of psychology is feebly ceding the territory of forgetting to the field of neuroscience, where investigations into the issue continue apace. In thinking about the concept of forgetting, the field should also consider whether it has anything interesting to say about why it occurs. **(p.336)**