Handbook of Research Methods in Human Memory

Hajime Otani, Bennett L. Schwartz

Measures of Forgetting

Publication details
Benjamin C. Storm
Published online on: 22 Oct 2018

How to cite: Benjamin C. Storm. 22 Oct 2018, Measures of Forgetting from: Handbook of Research Methods in Human Memory Routledge
Accessed on: 24 Sep 2019

PLEASE SCROLL DOWN FOR DOCUMENT

Full terms and conditions of use: https://www.routledgehandbooks.com/legal-notices/terms

This Document PDF may be used for research, teaching and private study purposes. Any substantial or systematic reproductions, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The publisher shall not be liable for an loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.
MEASURES OF FORGETTING

Benjamin C. Storm

That thoughts and experiences become inaccessible constitutes one of the most universally accepted observations of human memory and yet also one of its most enduring mysteries. How do we forget? Why do we forget? The present chapter aims to provide an overview of some of the most common methods used in the study of forgetting, focusing in particular on behavioral measures of forgetting in long-term memory. I will begin by discussing the ways in which forgetting is measured and then turn to outlining some of the most basic mechanisms thought to be responsible for forgetting and the various methodological approaches that have been used to study them.

Tulving defined forgetting as the “inability to recall something now that could be recalled on an earlier occasion” (1974, p. 74). Applying this definition to the context of a typical memory paradigm—consisting of an encoding phase, a retention interval, and a final test—forgetting can be operationalized as the impaired level of performance at test compared to what would have been observed at some earlier point in time, either at the point of encoding or at some point during the retention interval. To study forgetting, researchers often manipulate various aspects of this paradigm, such as the way information is encoded, the length of the retention interval, what happens during the retention interval, or the nature of the final test, hoping to use such manipulations to shine light on how and why forgetting occurs. Assuming equivalent levels of encoding, for example, if test performance in one condition differs from that of test performance in another condition, then the difference can be generally inferred to reflect a difference in forgetting.

In most studies, forgetting is measured as the difference in the proportion of items that are recalled in one condition (e.g., an experimental condition) compared to that of another condition (e.g., a baseline condition). The baseline condition provides a measure of how many items are successfully encoded and recalled at test independent of whatever variable is manipulated. If participants recall 4 out of 10 items (40%) in the experimental condition, for example, and 6 out of 10 items (60%) in the baseline condition, then the manipulation can be said to have caused a 20% forgetting effect. When measured in this way, a forgetting effect reflects a difference score between two conditions as well as the proportion of items from the entire distribution (i.e., all to-be-remembered items in the study list) that were recalled in the baseline condition but not in the experimental condition. Sometimes forgetting is measured in a slightly different way. Instead of calculating a difference score between two conditions, a proportion score is calculated between them. If participants recall 4 and 6 items in the experimental and baseline conditions, respectively, for example, then a forgetting effect can be measured as the difference between the two conditions divided by the total number of items recalled in the baseline condition (i.e., 2/6 or 33%). When measured in this way forgetting
Measures of Forgetting

reflects the proportion of items not recalled in the experimental condition that were recalled in the baseline condition.

Difference scores and proportion scores have their own sets of advantages and disadvantages, and either can be used as long as the decision is justified. Researchers should be careful though when comparing forgetting effects across individuals or between conditions, especially when there are large differences in baseline performance. In such situations the two measures can provide very different patterns of results, sometimes even showing a larger forgetting effect in one condition when using one measure and a larger forgetting effect in the other condition when using the other measure. If one group of participants recalls 15% of the items in the experimental condition and 25% of the items in the baseline condition, for example, and if another group recalls 45% of the items in the experimental condition and 65% of the items in the baseline condition—a pattern which would be expected if you were comparing groups of participants with different levels of overall memory ability—then the first group would exhibit a smaller forgetting effect compared to the second group when forgetting is calculated via difference scores (10% versus 20%) and a larger forgetting effect when forgetting is calculated via proportion scores (40% versus 31%).

Another complication in measuring forgetting has to do with the difference between the recall of a distribution of items and the recall of a single item. Most measurements of forgetting reflect impairments in memory for a distribution of items, with participants failing to recall some proportion of items that were studied or recalled in the baseline condition. Sometimes, however, researchers are interested in saying something about the strength of individual items within a distribution, such as the extent to which individual items lose memory strength across a retention interval or as a consequence of some experimental manipulation. It may be tempting to say that if a 20% forgetting effect is observed for an entire list of items that the average loss in memory strength of an individual item within that list would also be about 20%, but clearly this does not need to be the case. In fact, under some conditions a forgetting effect can be driven completely by the impairment of only a small subset of items within a distribution, and under such conditions it would be inappropriate to conclude that all items (or even that a majority of items) within the distribution were susceptible to forgetting (e.g., Kornell, Bjork, & Garcia, 2011; Storm, Friedman, Murayama, & Bjork, 2014). Most measures of forgetting reflect the average consequence of some manipulation on the recall of an entire distribution of items, a measure that cannot be used to infer the extent to which individual items within that distribution were affected by the manipulation.

Making matters even more complicated is the fact that test performance itself is sometimes not a valid indicator of whether a given item has been forgotten. Participants may refrain from overtly producing an item at test even though the item is covertly retrieved. This can occur for a variety of reasons but perhaps most frequently because participants are unsure of whether the item is the actual item they are attempting to recall. In episodic memory tasks, for example, participants are asked to retrieve items from an earlier study phase. If an item comes to mind the participant may be uncertain as to whether it is coming to mind because it was studied earlier or because it was generated accidentally from semantic memory. If a participant decides not to overtly output the item then such a behavior would contribute to the forgetting effect that is observed even if the item was (at least in some sense) successfully retrieved. Whether or not it would be appropriate to include such an item in the forgetting effect would depend on the type of forgetting that researchers are hoping to measure. In fact, some forgetting research is specifically interested in exploring the nature of what can and cannot be remembered at the time of test. In the remember/know procedure, for example, participants are shown items at test and asked to indicate whether they explicitly remember them being presented or simply know that they were presented, with impairments in the remembering or recollection component of the task presumably saying something different about the nature of a given forgetting effect than impairments in the knowing or familiarity component of the task (Donaldson, 1996; Knowlton & Squire, 1995; Tulving, 1985).
In studying forgetting, researchers implement measures that tap into the nature of the memory representations they are attempting to investigate. Every measure provides a somewhat different source of information. Impairments in recognition (Mandler, 1980), failures to remember to do something (Einstein & McDaniel, 1996), and failures to remember prior experiences of remembering (Schooler, Bendiksen, & Ambadar, 1997), to provide just a few examples, each say something different about the nature of what is being forgotten in a given circumstance. Moreover, it is important to measure forgetting in a way that is appropriate for the real-world or applied conditions to which one hopes to generalize. For example, there is a difference between the likelihood of a memory springing to mind involuntarily and the likelihood of it being retrieved given a specific retrieval cue (Berntsen, 2010). The difference is important because a measure of memory that provides participants with a specific cue may not say much about the probability of something being retrieved in real-world situations where such cues may not be available. An autobiographical experience might never be retrieved, for example, until a particular cue is encountered. In this way, a potentially useful measure of forgetting may not only be the extent to which a memory can be recovered given a particular cue, but the extent to which it will be recovered when not given a cue.

Another important consideration, especially from a theoretical perspective, is what Tulving and Pearlstone (1966) referred to as the distinction between availability and accessibility. When most people think about forgetting, they do not think about an item becoming temporarily inaccessible—rather, they think about something more akin to the permanent loss of storage, the idea that a memory is no longer available or represented in memory. The distinction between availability and accessibility poses a serious challenge to defining what it means for a memory to be forgotten. As is clear from past research, the underlying availability of an item—as reflected by its general fixture or embeddedness in memory—can vary substantially from its momentary state of accessibility (Bjork & Bjork, 1992; Estes, 1955). Changes in accessibility can occur despite little or no change in availability, and changes in availability can occur without concomitant changes in accessibility. As demonstrated in work on hypermnesia, for example, items that are not recallable on one test can recover spontaneously and become recallable on a subsequent test (Brown, 1923; Erdelyi, 1996; Payne, 1987). Such findings, along with many others, suggest that availability cannot be inferred directly from performance on any one test of memory.

To conclude that a memory is forgotten in some permanent sense requires evidence that it cannot be recalled given any possible cue or condition, a bar that is in practice far too high to reach. Given this limitation, many researchers have chosen to focus on studying forgetting in terms of impairments in accessibility. Although often appropriate, this decision does not change the fact that some permanent or enduring form of forgetting undoubtedly occurs, and that some theoretical questions require researchers to attempt to differentiate between availability and accessibility in their studies. Even if a direct measure of availability cannot be achieved, researchers can still try to approximate relative differences in something like availability by employing measures less likely to be susceptible to momentary changes in performance, such as by looking at savings in relearning or by measuring performance across a variety of contexts.

Finally, it is important to consider the assumptions and metaphors brought to bear when thinking about memory (Roediger, 1980). The language we use when we say that an item has been encoded, stored, or retrieved, for example, affects the theoretical and methodological decisions that we make in subtle but powerful ways. We act as if an item, once encoded, is stored in some way until it is later retrieved, and that when taking a test a participant is retrieving the same item that had been originally encoded. As has been known for some time, however, memory reflects a process of construction, not reduplication (Bartlett, 1932). It does not simply store the past—it prepares us to behave more adaptively in the future by providing the building blocks needed to reconstruct the past, imagine the future, and inform our various judgments and decisions (Schacter, 2001, 2012). From this perspective, simply trying to measure and dissociate things like availability
Measures of Forgetting

and accessibility may not be sufficient. When a memory is retrieved, such a memory will always, to at least some extent, be different from whatever learning took place during study. What is being measured, therefore, in most studies of forgetting, is not the reduced accessibility of an item in memory (let alone its availability), but the extent to which a particular experience at one point in time fails to influence a particular behavior or experience at another point in time. Ultimately, the proper definition of forgetting, and thus the proper method of measuring forgetting, will always depend on what a researcher is attempting to understand. Whatever the aim, researchers would be well served to clearly define what they mean by forgetting and to operationalize their variables and measures accordingly.

Methods for Studying the Mechanisms of Forgetting

The methods used in any area of research must be considered within the context of the questions and theories they are designed to investigate. Most of the research on forgetting has been designed with the specific goal of uncovering the mechanisms by which experiences once remembered become forgotten. In organizing the rest of the present chapter, therefore, it felt only natural to do so by outlining a few of the most prominent mechanisms that have been argued to cause forgetting and to discuss the methodological approaches that have been used to study them. Each section will also include examples of complications and challenges encountered in these contexts. The coverage is in no way meant to be exhaustive, or even fairly representative, of the entire area of research on forgetting. Hopefully the discussion will serve as a useful starting point, however, for measuring and interpreting the effects of forgetting with implications ultimately applicable to any number of research situations.

Time and Disuse

One of the oldest and most intuitive theories of forgetting assumes that impairments of memory are caused by the passing of time. The idea has a long history, dating back to Thorndike’s (1914) Law of Disuse. According to this idea, it is because of disuse that memories become weakened to the point that they are no longer recoverable. Although alluring in its simplicity, the Law of Disuse, at least in its strongest form and in the context of long-term memory, has been roundly discredited (Bjork & Bjork, 1992; McGeoch, 1932). Researchers now know that forgetting occurs with time but that time itself is not necessarily responsible for forgetting. Indeed, the relationship between time and forgetting is often not as strong as one would expect and when there is a relationship such a relationship can often be explained by factors independent of time. Nevertheless, the possibility that time plays at least some role in causing forgetting remains an empirical question. The paragraphs below briefly present a few of the ways in which researchers have sought to separate the influence of time from some of the factors with which it tends to be confounded.

One particularly famous example of research exploring the effect of time on forgetting, specifically in the context of short-term memory, is the work of Brown (1958) and Peterson and Peterson (1959). Participants learned short lists of items and then counted backwards for a certain amount of time before being tested. By counting backwards participants were prevented from engaging in any type of rehearsal, and the stimuli were designed in such a way that the counting task would be unlikely to cause associative interference. The classic finding was a forgetting curve in which the items were highly accessible immediately after learning and then quickly lost across the up to 18-second retention interval that followed. As subsequent research has shown, however, this forgetting function was due more to a build-up in proactive interference than it was to decay. When only the first trial was examined, for example, or when participants were given a list of items not susceptible to proactive interference, the forgetting effect was largely eliminated (Keppel & Underwood,
1962; Wickens, 1973). One of the many lessons learned from this research is that there can be an important trade-off between power gained from having many observations and validity lost from having earlier trials have the potential to influence performance on later trials.

Another way in which researchers have attempted to isolate the effects of time on forgetting is by measuring the consequences of a period of mental quietude—that is, a period of time in which little or nothing occurs in a person’s mind. Unfortunately, this too has proven difficult to operationalize. Even when participants are sitting quietly and not learning anything new, they are still engaging in a collection of cognitive processes that have the potential to affect memory, whether it is by rehearsing earlier learned information or by engaging in any other type of internally focused cognitive activity (Andrews-Hanna, 2012; Smallwood & Schooler, 2006). One might argue that sleep provides a closer approximation of mental quietude, as at least it involves a state of unconsciousness. Indeed, as shown by Jenkins and Dallenbach (1924), and by many others since, participants who sleep during a retention interval tend to forget less of what they learned prior to sleep than participants who remain awake. To be sure, such findings are interesting, but they do not provide compelling evidence that little or no forgetting occurs as a function of time. For one, sleep is decisively not a period of mental quietude, and part of what sleep may do is to help solidify or stabilize recent learning to make it less susceptible to forgetting (Walker & Stockgold, 2006).

Even if a satisfactory method of isolating the effect of time on forgetting fails to ever emerge, this does not mean that time should be ignored as a variable of interest in the study of forgetting. First, time can serve as a proxy for other variables that are more difficult to measure and that tend to be correlated with time. Second, from a practical standpoint, researchers sometimes simply want to know whether a given item or set of information will be recallable or continue to affect performance after a delay. When investigating the effect of some training or educational intervention, for example, one might want to measure the extent to which some new learning is forgotten across the course of a person’s training or career, in which case time is of interest even if not being used as a theoretical explanation.

Finally, it is not uncommon for researchers to report evidence of equivalent or even improved levels of recall after a long delay compared to a short delay—showing, for example, that performance is better after 24 hours than after 20 minutes. Although researchers may be tempted to conclude that forgetting was not observed, or that some form consolidation process must have taken place during the delay, a much more mundane explanation can often be invoked. Specifically, for what are likely a variety of reasons, participants who sign up for a two-part study and who come back for the second part of a two-part study may simply do better on memory tasks than participants who do not, creating a selection bias confounded with the delay manipulation. Researchers can safeguard against this possibility by requiring all participants, regardless of condition, to sign up for the same delayed session and then only include data from participants who come back for that session. Such control can be onerous, especially when participant hours are at a premium, but it is absolutely necessary if comparisons across delay are of practical or theoretical interest. Fortunately or unfortunately, depending on one’s perspective, this issue has been relatively inconsequential to date because most memory studies have focused on the use of brief, hour-long experimental sessions.

**Interference**

According to interference theory, as a cue becomes associated to multiple responses there is competition between those responses. By learning new responses to a given cue, therefore, or by heightening the accessibility of a select set of responses associated to that cue, the recall of other responses becomes impaired (Crowder, 1976; McGeoch, 1942; Postman & Underwood, 1973). Over the years, researchers have developed many methods to observe and illuminate the ways in which interference causes forgetting. In one particularly popular paradigm, participants study a series of
paired associates, such as “cloud-dog.” The first word is referred to as the stimulus or cue, and the second word is referred to as the response or target. Using this paradigm, researchers are able to manipulate the nature of the cue-response pairs as well as the relationships between the pairs. In one design, for example, participants might study a new response, such as “phone” as an associate to the cue “cloud.” At test, this new response is said to retroactively interfere with the recall of the old response, whereas the old response is said to proactively interfere with the recall of the new response. Such interference effects can only be measured with appropriate baselines. To measure retroactive interference, for example, recall performance for old responses must be compared to performance when new responses are not learned or when new responses are learned but they are not related to the original cues.

In studying associative interference, different baselines and measures are critical for answering different empirical and theoretical questions (Postman, 1971). One variation that has received considerable attention in the literature is that of the final test. In one version, for example, participants are given each cue and asked to recall a particular response, such as the first response that was encountered during learning. In other versions, participants are given each cue and asked to retrieve either the first response that comes to mind (Modified Free Recall) or all of the responses that were learned (Modified Modified Free Recall). Presumably, the levels of performance observed using these and many other types of tests provide unique insight into the relative accessibility of information in memory as well as into the dynamics of how interference causes forgetting. A test that allows for multiple responses, for example, is likely to provide a better measure of the extent to which a particular item can be remembered than a test that allows participants to make only one response. The latter measure, however, may be better suited for assessing the relative accessibility of different items in memory as well as the extent to which one item dominates other items in response to a particular cue.

Research on interference has not been limited to the type of paired-associate paradigm described above. Indeed, interference effects can be observed in just about every measure of memory. If participants are given a list of words to learn, for example, and asked to recall them at test, the words within that list will interfere with each other in ways that impact performance. The more items there are on a list, the less likely any one individual item will be recalled, and the items that are learned at the beginning and end of the list will be better recalled than those learned in the middle (Murdock, 1962). Making matters even more complicated is the fact that recall reflects a moving target. More specifically, retrieval acts as a self-limiting process in that the very act of retrieving one item alters the likelihood of being able to retrieve another item, a phenomenon known as output interference (Roediger, 1978).

Though interesting in its own right, interference poses serious challenges to interpreting the role of other mechanisms in causing forgetting. It is often unclear, for example, whether forgetting can be attributed to an investigator’s manipulation of interest—such as the way in which items are encoded or what happens during a retention interval—or to interference dynamics confounded with such manipulations. A researcher comparing the retention of emotional versus non-emotional information, for example, assuming equivalent levels of learning at encoding, might interpret differences in test performance as reflecting differences in the extent to which such information is forgotten across the retention interval. If participants recall emotional items first, however, as would likely be the case in situations of free recall, then any difference in performance could be at least partly attributed to output interference at test and not to the relative forgetting rates of emotional and non-emotional items.

Finally, it is worth noting a problem associated with a measure of interference that is sometimes employed, particularly when baseline levels of performance are not available. Specifically, researchers may measure the rate of intrusions at test from non-target items. Such a measure can provide a useful approximation of the degree of interference, but not always. To understand why this is the
case one simply needs to think about the many situations in which the correlation between interference and intrusions breaks down. Participants are less likely to make intrusion errors, for example, when they know that a given item is incorrect. Knowing that an item is incorrect and thus not outputting it at test, however, does not mean that it did not block or otherwise impede the recall of other items. In fact its covert retrieval would seem sufficient to cause output interference. To be sure, intrusion rates can and do provide important information about the relative accessibility of different information in memory, but the amount of intrusions should not be considered direct evidence of the degree of interference. In particular, the absence of intrusions should not be considered evidence of the absence of interference.

Cues and Context

Another important source of forgetting relates to the set of cues or contextual conditions available at the time of test (McGeoch, 1932; Estes, 1955; Tulving & Thomson, 1973). Whether a particular item can or cannot be retrieved is often determined by whether an appropriate cue or set of cues is provided. An item highly accessible in one context, for example, might become entirely inaccessible in another context, and as a particular cue becomes overloaded with potential responses the likelihood of recalling any one response becomes progressively impaired (Watkins & Watkins, 1975). In this way, changes in cues and context (or their relative effectiveness) can be considered a mechanism for causing forgetting.

One popular approach to studying the role of cues and context in forgetting has been through iterations of the encoding/retrieval paradigm (Roediger & Guynn, 1996; Tulving, 1983). In this paradigm, a 2 × 2 factorial design is created by crossing two encoding conditions (A and B) and two retrieval conditions (A′ and B′), allowing researchers to examine whether performance is better when there is a match between encoding and retrieval (A-A′ and B-B′) than when there is a mismatch (A-B′ and B-A′). The use of the encoding/retrieval paradigm has been quite productive. Researchers have manipulated a variety of factors, from verbal context to environmental context to mood and state-dependent context, all of which have been shown to impact the items that can and cannot be recalled (Eich & Metcalfe, 1989; Godden & Baddeley, 1975; Smith, Glenberg, & Bjork, 1978). As expected, participants tend to perform worse in the mismatched conditions than in the matched conditions, presumably because they do not have the cues necessary to support successful recall (cf. Nairne, 2002).

Although changes in context can certainly cause forgetting (at least as reflected by impaired levels of accessibility), whether and to what degree such forgetting is observed often depends on methodology. As Smith and Vela (2001) found in their meta-analysis, context-based forgetting effects are reduced or even eliminated when participants use non-context-based cues at either encoding (i.e., overshadowing) or test (i.e., outshining). In the case of outshining, for example, if the cues given to participants at test are sufficient to allow an item to be recalled even without the help of a given context, then the deleterious effect of not having that context available will be diminished. In other words, changes in context should only cause forgetting if the original context is needed to support performance. Such dynamics explain why context effects are typically smaller on tests of cued recall and item recognition than they are on tests of free recall, and why students taking a final exam in a different classroom should not be overly concerned about their performance being negatively affected compared to students taking the final exam in the same classroom. In such situations the cues and questions themselves are likely to outshine any potential effect of learning context. Another important methodological consideration is the participant’s tendency or ability to mentally reinstate the original context. A researcher might remove certain cues at test but if participants are able to mentally bring those cues back then the effect of their removal is mitigated. Given these observations, researchers hoping to observe context-dependent effects of forgetting should (1) design their paradigm in a way that encourages participants to interact with contextual cues at the time of
Measures of Forgetting

encoding, (2) employ a final test for which performance depends on the availability of the contextual cues that were present at the time of encoding, and (3) prevent participants from intentionally or spontaneously reinstating the original context at test.

Finally, cues are important to consider not only as they relate to the memory being targeted (e.g., the strength of the cue-response association), but in the way they influence the strategies and metacognitive judgments people employ in their retrieval efforts. When experiencing a tip-of-the-tongue state, for example, a person may not be able to retrieve a particular item in memory, such as a person’s name or a low-frequency word, but they may nevertheless know (or think they know) on the basis of the cues available that the item is indeed somewhere in memory waiting to be retrieved (Brown & McNeill, 1966; Schwartz, 2002). In such cases, cues create a feeling of knowing that contributes to a person’s experience of forgetting. Although such feelings are often correct, they can also be very wrong, leading people to think that they know something even if they do not. Such dynamics are important not only for studying tip-of-the-tongue states, but more broadly for studying the phenomenology of forgetting and for identifying the conditions under which people are likely to anticipate forgetting or persist in their efforts to overcome states of forgetting (e.g., Finn, 2008; Schwartz & Metcalfe, 2011). Perhaps even more generally, however, such dynamics suggest that forgetting cannot be inferred directly on the basis of a person’s subjective experience (Koriat, 2000). Just because a participant reports forgetting does not mean that something has been forgotten. Of course, the interpretation of this consideration depends on what one intends to measure when studying forgetting. The experience of forgetting may be just as real and consequential to a person who feels like they are failing to remember as it is to a person who is actually failing to remember.

Inhibition and Goal-Directed Forgetting

A significant portion of the research on forgetting over the past few decades has focused on the potential role of inhibition in causing what can be referred to as goal-directed forgetting (Bjork, Bjork, & Anderson, 1998; Bjork, 1989; Storm, 2011). At the core of this area of research is the idea that in order to function adaptively there needs to be some means of overcoming the accessibility of outdated, unwanted, or contextually inappropriate information in memory. By inhibiting outdated information, for example, people place themselves in a better position to learn and remember new and more useful information. Such inhibition is presumed to come at a cost, however, in that the inhibited information remains inaccessible at a later time even when it becomes the target of recall. Several paradigms have been argued to provide evidence of this sort of forgetting, three of which will be discussed below.

In the list-method paradigm of directed forgetting (Bjork, 1970; MacLeod, 1998), participants learn an initial list of items before learning a second list of items. Critically, after learning the first list, participants are told that they should either continue to remember the first list (remember condition) or that they should forget it (forget condition). The prompt to forget can be accomplished in many ways. Participants might be told that the list was presented as practice, for example, or that the researcher made a mistake and presented it by accident. In any case, participants are led to believe that they no longer need to remember the first list, a realization that has two important consequences: The first list becomes less recallable than it would have been otherwise (the costs of directed forgetting), and the second list becomes more recallable than it would have been otherwise (the benefits of directed forgetting). According to the inhibition account, the first list is suppressed during the learning of the second list in order to facilitate the learning and remembering of the second list (for a recent review of the many methodological considerations to be taken into account when studying list-method directed forgetting, see Sahakyan, Delaney, Foster, & Abushanab, 2013).

A somewhat different type of intentional forgetting has been studied using the Think/No-Think (TNT) paradigm (Anderson & Green, 2001). In this paradigm, participants first learn a list of
cue–response pairs, such as “ordeal-roach.” Then, during one version of the TNT phase, the cues of some of the pairs are presented in green, whereas the cues of other pairs are presented in red. When presented in green, participants are told to think of the associated response (Think items). When presented in red, participants are told to not think of the associated response, and in fact to not even let the response enter consciousness (No–Think items). A third set of cues are not presented at all during the TNT phase (Baseline items). After many rounds of trials, a final test is administered in which participants are asked to recall the previously learned responses. Forgetting is observed when No–Think items are recalled less well than Baseline items, a result which suggests that the attempts to suppress caused forgetting.

The white bear paradigm developed by Wegner and colleagues provides an interesting juxtaposition to the work on TNT, showing to-be-suppressed information to become more accessible, not less accessible, following suppression (Wegner, Schneider, Carter, & White, 1987). Several differences appear relevant to explain this apparent contradiction. Unlike in the white bear paradigm, for example, TNT participants are not instructed to monitor their suppression attempts, nor are they required to suppress an item for more than a few seconds at a time. Such factors suggest that there are conditions in which suppression is more likely to succeed than others and that researchers hoping to measure suppression-induced forgetting should keep such factors in mind. Indeed, researchers have sometimes reported difficulty in replicating the TNT effect, and such difficulties may be attributed at least in part to methodological factors much more subtle than those listed above, such as the wording of the instructions, participant compliance, and experimental fatigue (Anderson & Huddleston, 2011). In many studies, for example, participants may occasionally check their memory in order to ascertain whether a given to-be-suppressed item can still be recovered, a behavior that would presumably reduce the effectiveness of the suppression manipulation. To prevent this behavior, Anderson and Huddleston have encouraged researchers to eliminate any mention of memory in their instructions or consent forms, and to emphasize instead that the study is about the control of attention. The use of a post-experimental questionnaire can also be useful for identifying participants who engage in this type of behavior or who might otherwise fail to comply with the suppression instructions for any other reason. Ultimately, for TNT effects to be observed, it is crucial for participants to buy into the suppression instructions and to engage fully in their suppression attempts.

A final example of goal-directed forgetting is that of retrieval-induced forgetting (Anderson, Bjork, & Bjork, 1994). Retrieval-induced forgetting is different than directed forgetting and TNT in that it is generally not believed to be the consequence of a conscious or intentional decision to suppress non-target information. In the retrieval-practice paradigm, participants study a list of category-exemplar pairs (e.g., fruit–lemon, drinks; rum, fruit–peach, drinks: vodka) before receiving retrieval practice for half of the exemplars from half of the categories (e.g., fruit–le______). Retrieval practice creates three types of items: Rp+ items (practiced exemplars from practiced categories; lemon), Rp– items (non-practiced exemplars from practiced categories; peach), and Nrp items (exemplars from non-practiced categories; rum, vodka). Finally, participants are tested on all of the studied items, and retrieval-induced forgetting is observed when Rp– items are recalled less well than Nrp items. According to the inhibition account, Rp– items are inhibited during retrieval practice because they cause competition with the recall of Rp+ items, and it is the consequence of this inhibition that persists until the time of the final test (Anderson, 2003; Storm & Levy, 2012). Interestingly, a number of factors—such as the relationships between items, the form of retrieval practice, and the nature of the final test—have been shown to influence whether and to what degree retrieval-induced forgetting is observed (for reviews, see Murayama, Miyatsu, Buchli, & Storm, 2014; Storm et al., 2015).

Researchers hoping to use these paradigms to make inferences about inhibition and goal-directed forgetting should do so with caution and be sure to distinguish between the empirical effect they observe and the theoretical mechanism thought to underlie it. In the study of retrieval-induced
Measures of Forgetting

forgetting, for example, just because Rp-items are recalled less well than Nrp items does not mean that those items were inhibited during retrieval practice. Indeed, many other mechanisms have been argued to explain retrieval-induced forgetting, such as changes in context and associative interference (Jonker, Seli, & MacLeod, 2013; Raaijmakers & Jakab, 2013). Although there may be no perfect method for isolating the role of inhibition in causing forgetting, certain approaches seem to work better than others, such as using item-specific cues at test instead of more general category cues (Schilling, Storm, & Anderson, 2014; Storm & White, 2010). Ultimately, when making conclusions about the possible role of inhibition, it is imperative for researchers to argue or show that a given effect cannot be explained by mechanisms other than inhibition. Just how high the bar stands for providing this sort of evidence, however, and even what such a bar would look like, remains a topic of ongoing debate.

Disrupted Consolidation and Reconsolidation

Factors such as inhibition, interference, and context change are assumed to cause forgetting by disrupting or impeding performance. That is, they are assumed to cause temporary reductions in accessibility that can be reversed or eliminated. In contrast, factors related to consolidation and reconsolidation are assumed to cause more permanent changes in memory, not by impairing the accessibility of a given memory, but by altering or modifying the availability of the representations underlying that memory.

Müller and Pilzecker (1900) provided one of the earliest theories of forgetting in their perseveration-consolidation hypothesis. According to the hypothesis, activity in the brain perseverates after initial exposure or training to consolidate what has been learned into a more permanent form. New learning that occurs after initial learning, however, has the potential to disrupt the consolidation process and cause the initial learning to be forgotten, not only in the sense of making it less accessible, but by diminishing its long-term integrity or stability (McGaugh, 2000; Wixted, 2004). Researchers over the years have employed a variety of methods to explore the ways in which the disruption of consolidation can cause forgetting. In one behavioral approach, the delay between initial learning and later learning is manipulated. Presumably, when the delay is relatively short, initial learning has less time to consolidate and should therefore be more susceptible to disruptions in consolidation than when the delay is relatively long. Other approaches have relied on patients with damage to areas of the brain responsible for consolidation or on the use of pharmacological interventions. It has been shown, for example, that drugs that prevent the consolidation of new learning while not disrupting the consolidation of earlier learning can reduce the extent to which earlier learning is forgotten.

The role of consolidation in memory has received increased attention more recently in the study of reconsolidation. According to the theory of reconsolidation, when memories are reacti-

vated they become labile and susceptible to being overwritten or modified (Dudai, 2004; Misanin, Miller, & Lewis, 1968; Nader, Schafe, & Le Doux, 2000), a mechanism that may account for a variety of memory phenomena such as misinformation effects and the modifying consequences of retrieval (Loftus, 2005). The upwelling of interest in reconsolidation has increased substantially in the last couple decades, with numerous observations of reconsolidation-induced disruptions in fear memories of animals (Nader & Hardt, 2009) and humans (Schiller et al., 2010). More recently, observations have been made in other contexts as well, such as with procedural memories (Walker, Brakefield, Hobson, & Stickhold, 2003) and declarative memories (Hupbach, Gomez, Hardt, & Nadel, 2007).

Research on reconsolidation continues to emerge. A challenge to providing strong empirical evidence in support of reconsolidation, however, is that it is difficult to rule out other, non-reconsolidation-based mechanisms. In a typical paradigm used to study reconsolidation in episodic
memory (Hupbach, Gomez, & Nadel, 2013), for example, a memory is reactivated (or not reactivated) before participants are presented with some new information that is related or in conflict with that memory. At final test, forgetting due to reconsolidation is inferred to the extent that the reactivated memory is impaired as a consequence of the presentation of the new information. Importantly, it must also be shown that forgetting is not observed if the memory is not reactivated or if the new information is presented outside of some reactivation window. One difficulty that arises, however, which was discussed earlier in the chapter, is that it is difficult to show conclusively that forgetting reflects a permanent loss in availability, as would be the case if forgetting was being caused by reconsolidation. Impaired performance might be attributed to source confusion, for example, or to changes in context or susceptibility to associative interference. If researchers hope to make inferences about the role of reconsolidation in causing forgetting they should put methods in place capable of addressing these counter-explanations.

Finally, researchers interested in studying the role of consolidation and reconsolidation in forgetting should be careful to define what they mean by the term consolidation (Wixted & Cai, 2013). Is it referring to the extent to which a memory becomes less dependent on structures in the hippocampus and medial temporal lobe (systems consolidation)? Or is it referring to the extent to which a memory trace becomes stabilized (cellular consolidation)? Is it referring to the enhanced creation or integration of initial learning? Or is it referring to a reduction in the susceptibility of initial learning to the destructive forces of new learning? By clearly conceptualizing consolidation, researchers place themselves in a better position to select the methods and measures most appropriate for testing their research question and to have the greatest possible impact on the literature.

Concluding Comment

As may be clear from the sections above, memory researchers love a good paradigm. Whether it is using the paired-associate paradigm, the encoding/retrieval paradigm, or the retrieval-practice paradigm, much of the published research on forgetting has focused on the thorough and often very thoughtful investigation of particular phenomena using particular paradigms. This focus is quite understandable, as the use of well-trodden tasks and measures provides a number of methodological advantages, from allowing researchers to more easily build upon prior work to facilitating the use of meta-analysis. Such virtues, however, come at a cost. Paradigms are developed to test hypotheses, but they can quickly become the targets of research themselves (Hintzman, 2011). Researchers should take advantage of existing paradigms, but they should not lose sight of the more general questions their research is intended to answer (e.g., questions concerning the fundamental nature of how and why people forget), which will often require significant divergences from the tried and true tasks and measures of the past.

References

Measures of Forgetting


Measures of Forgetting


