When Remembering Causes Forgetting: Retrieval-Induced Forgetting as Recovery Failure

Karl-Heinz Bäuml, Martina Zellner, and Roman Vilimek
Regensburg University

Retrieval practice on a subset of previously learned material can cause forgetting of the unpracticed material and make it inaccessible to consciousness. Such inaccessibility may arise because the material is no longer sampled from the set of to-be-recalled items, or, though sampled, its representation is not complete enough to be recovered into consciousness. In 2 experiments, it was examined whether retrieval-induced forgetting reflects a sampling or recovery failure by studying the time course of cued recall in this type of situation. Although retrieval practice reduced recall totals of the unpracticed items, in both experiments, the forgetting was not accompanied by an effect on the items' response latencies. This pattern of results is consistent with the view that inhibited items are successfully sampled but, because of a reduction in their activation level, do not exceed the recovery threshold.

Keywords: episodic forgetting, retrieval-induced forgetting, sampling, recovery

A form of episodic forgetting, which has been extensively studied in recent years, is retrieval-induced forgetting (for reviews, see Anderson, 2003; Levy & Anderson, 2002). Retrieval-induced forgetting can be demonstrated by letting participants learn items from different semantic categories. After a distractor task, participants practice on half of the items from half of the categories. The word stem of the to-be-practiced items is provided, and participants are asked to name the item that belongs to the cue. After another distractor phase, participants are asked to recall all of the previously presented material, given the category names as retrieval cues. The main finding in this experiment is that, relative to the unpracticed items from the unpracticed categories—which serve as a baseline—retrieval practice improves recall of the practiced material and impairs recall of the unpracticed material.

Retrieval-induced forgetting is a recall-specific effect (Anderson, Bjork, & Bjork, 2000; Bäuml, 2002; Ciranni & Shimamura, 1999) and has been observed with a wide variety of stimulus classes. It has been found with verbal and visual material (Ciranni & Shimamura, 1999) and has proven relevant in a number of settings such as fact learning (e.g., Anderson & Bell, 2001), eyewitness memory (e.g., Shaw, Bjork, & Handal, 1995), false memories (e.g., Bäuml & Kuhbandner, 2003), memory for percep-

tual experiences (e.g., Schooler, Fiore, & Brandimonte, 1997), and social cognition (e.g., Dunn & Spellman, 2003). It occurs in recall and recognition (e.g., Hicks & Starns, 2004) and can be found in both episodic and semantic memory (e.g., Johnson & Anderson, 2004). It has mostly been observed in young adults but is present in young children (Zellner & Bäuml, 2005) and older adults (Moulin et al., 2002) as well. Taken together, retrieval-induced forgetting appears as a very robust and general phenomenon.

Accounts of retrieval-induced forgetting are often based on the proposal that this type of forgetting is caused by inhibition (Anderson, Bjork, & Bjork, 1994; Anderson & Spellman, 1995; for a noninhibitory account, see Williams & Zacks, 2001). This proposal rests on the assumption that items that share a common cue compete for conscious recall if that cue is provided. If one of the items then is practiced, the other items sharing the same cue are assumed to interfere and, to guarantee a successful recovery of the to-be-practiced item, need to be inhibited. This inhibitory account of retrieval-induced forgetting contrasts with theoretical accounts of other forms of episodic forgetting, like interference, in which the forgetting is supposed to be the result of enhanced competition (Raaijmakers & Shiffrin, 1981) or part-list cuing, in which it has been suggested that the forgetting is caused by retrieval competition (Rundus, 1973) or strategy disruption (Basden & Basden, 1995).

Although there is broad agreement in the literature that the goal of inhibition in retrieval-induced forgetting is to reduce the interference potential of competing material, it is still largely unclear what the nature of the inhibitory mechanism is. A priori, there are several ways in which inhibition might reduce the interference potential of competing items. One way would be that inhibition initiates an unbinding process and disconnects interfering items from the common cue, a proposal repeatedly suggested in the literature on episodic forgetting (Geiselman, Bjork, & Fishman, 1983; Postman, Stark, & Fraser, 1968). Another way would be that inhibition reduces the speed of processing of interfering items and thus induces a slowing in retrieval of the interfering material, a

proposal entertained in the negative priming literature (Conway, Tuholski, Shisler, & Engle, 1999; Tipper & Cranston, 1985). A third way would be that inhibition reduces the level of activation for a given item and thus prevents the item from achieving the recovery threshold, an hypothesis suggested in the retrieval-induced forgetting literature itself (Anderson, 2003; Veling & van Knippenberg, 2004). All three types of inhibitory mechanisms—unbinding, retrieval slowing, and reduction in activation level—would reduce the interference potential of competing material and thus increase recall chances of target material.

A common feature of the three inhibitory mechanisms is that they make inhibited items inaccessible to consciousness. Following two-stage models of recall, in which a sampling and a recovery stage of recall are incorporated (described below), there are at least two reasons why an inhibited item may become inaccessible for conscious recall. The one reason is that the item may fail to be sampled from the set of to-be-recalled items (sampling failure). The other reason is that, although sampled, the item's representation may no longer be complete enough to exceed the recovery threshold (recovery failure), thus mimicking the tip-of-the-tongue phenomenon in which an item may be sampled but cannot be recovered (Brown & McNeill, 1966). As explained below, both unbinding and retrieval slowing represent sampling accounts of retrieval-induced forgetting, whereas the reduction-in-activation-level hypothesis represents a recovery account.

The theoretical interpretation of retrieval-induced forgetting obviously depends on knowing which of the two properties of recall is affected. Not much is yet known about the issue. Fortunately, however, sampling-based and recovery-based accounts of retrieval-induced forgetting can be distinguished experimentally by examining items' response latencies. Pursuing this goal, the analyses presented in this article focus on response latency in retrieval-induced forgetting. Before considering these results, response latencies, sampling/recovery models of recall, and the role of these models for theoretical interpretations of forgetting in human memory are first described in more detail.

Response Latencies and Sampling/Recovery Models of Recall

When studying episodic forgetting in tests of free and cued recall, researchers typically focus on recall total, that is, how many items have been recalled during some brief period of time, regardless of the speed with which those items have been retrieved. Analysis of recall total would indeed be sufficient if recall total was strongly correlated with speed of retrieval so that, for instance, a decrease in recall total was always accompanied by an increase in latency, and vice versa. Results from a recent series of experiments by Rohrer, Wixted, and colleagues (Rohrer, 1996; Rohrer, Salmon, Wixted, & Paulsen, 1999; Rohrer & Wixted, 1994; Rohrer, Wixted, Salmon, & Butters, 1995; Wixted, Ghadisha, & Vera, 1997; Wixted & Rohrer, 1994), however, suggest otherwise. These researchers demonstrated that recall total and response latency are empirically independent, with some experimental manipulations affecting recall total but not response latency and others affecting response latency but not recall total. Response latencies thus can add something to the understanding of human memory and thus may also add something to the understanding of episodic forgetting.

Analysis of response latencies over a wide range of experiments has revealed that participants typically remember many items early in the recall period and relatively few items later in the recall period. Consistently, response latencies have been found to be well described by means of an exponential,

$$r(t) = (N/\tau)e^{-t/\tau},$$

where $r(t)$ represents the number of items recalled at time t (or, in practice, at a particular time interval t), N represents asymptotic recall (the estimated number of items that could be produced given unlimited time), and τ represents the mean response latency of those N items (Bousfield & Sedgewick, 1944).[1]

A very prominent account of the exponential form of the latency function is provided by the random search model (McGill, 1963). According to this model, a certain retrieval cue delimits a mental search set, and exemplars from this set are then randomly sampled one at a time, at a constant rate. Each sampled exemplar is immediately recognized as either a not-yet-sampled item (and then subject to the recovery stage) or a previously sampled exemplar (and then ignored). Then the item is replaced. Because of the increase in resampling of already retrieved items over the recall period, a gradual decline in recall arises with exponentially declining output rates. Although the random search model is an oversimplification of retrieval (Herrmann & Pearle, 1981; Morrison, 1979; Vorberg & Ulrich, 1987), it has proven an extremely useful and robust account of response latencies (for a review, see Wixted & Rohrer, 1994).

One of the most intriguing predictions of the random search model is that mean latency (tau [τ]) depends linearly on both the size of the mental search set and speed of processing (McGill, 1963). Thus, mean latency should double if the number of items in the mental search set doubles and should be reduced by half if the search set size is reduced by half. Mean latency should also increase if a secondary task is introduced at test; a secondary task should reduce speed of processing and thus slow the sampling process. Rohrer and Wixted (1994) examined how response latencies depend on search set size. They let participants study lists of three, six, and nine items that were presented once every 2 s. As predicted, mean latency (τ) increased with the search set size and roughly doubled (tripled) if the number of items in the mental search set doubled (tripled). Rohrer et al. (1995) also examined how retrieval was affected while performing a concurrent secondary task. The secondary task resulted in a slower mean latency (τ), indicating that mean latency depends directly on processing speed.

A third, somewhat counterintuitive prediction of the random search model has also been found to be valid. Because, according to the random search model, response latency (τ) depends only on search set size and processing speed, the absolute strength of the items in a search set—as, for instance, operationalized through item study time or number of study trials—should not affect mean

---

[1] In the current article, we describe response latencies using noncumulative latency functions, where $r(t)$ is defined as the number of items recalled at a particular time interval t. This contrasts with the cumulative form of latency functions, where $r(t)$ is defined as the number of items recalled by time t. Cumulative and noncumulative functions divulge the same information about the rate of retrieval; however, the noncumulative form has the advantage that successive data points do not depend on each other, which is important for statistical testing.

latency. Rohrer and Wixted (1994) let participants study a list of items at rates of 1, 2, or 4 s per word. As expected, longer study times strongly increased recall totals. As predicted, however, the variation in study time did not influence mean latency (for related results, see Rohrer, 1996; Wixted et al., 1997). Thus, although in general more items are recalled from a set of strong items than an equally sized set of weak items, the mean latencies (τ) of the two types of items are the same. All of these findings converge on the view that mean latency depends on search set size and processing speed but not on absolute item strength.

These results are consistent with a two-stage model of recall (Rohrer, 1996). According to the model, in the first stage, items are sampled from a set of to-be-recalled items according to a relative-strength rule. This rule says that the probability of sampling an item depends on the size of the mental search set and equals the item's absolute strength divided by the sum of all absolute item strengths. This rule together with processing speed determines the item's response latency. In the second stage, a sampled item is recovered into consciousness if its absolute strength exceeds some threshold. Stronger items with fairly complete memory representations will be recovered, weaker items with less complete representations will not. Thus, whereas an item will eventually be sampled, it may not be recovered. Similar two-stage conceptualizations of recall can be found in several computational models, like, for instance, variants of the search-of-associative-memory model (Gronlund & Shiffrin, 1986; Mensink & Raaijmakers, 1988; Raaijmakers & Shiffrin, 1981).

The conceptualization of recall as a two-stage process, in which a sampling and a recovery stage are incorporated, is consistent with a number of findings in the memory literature (Brown & McNeill, 1966; Raaijmakers & Shiffrin, 1980), and it is consistent with the observed independence of recall total and response latency (Rohrer, 1996; Rohrer et al., 1999). Following this conceptualization, an item's response latency is a measure of search set size and processing speed and thus provides a window into the sampling stage of recall, whereas an item's probability of being recalled is a measure of its (absolute) strength and provides a window into the recovery stage of recall. Analyzing recall total and response latency in concert thus may provide unique insight into the role of sampling and recovery processes in memory retrieval.

Sampling Accounts of Episodic Forgetting

Sampling/recovery models of recall can help to improve the understanding of episodic forgetting. In fact, if an episode is no longer recallable, does the forgetting occur because the episode is no longer sampled during retrieval, or does it occur because the episode is sampled but, as a result of a less complete memory representation, can no longer exceed the recovery threshold? The interpretation of forgetting depends on knowing whether the sampling stage or the recovery stage of recall is affected. The literature suggests that at least some forms of episodic forgetting reflect problems at the sampling stage of recall.

Wixted and Rohrer (1993) analyzed recall totals and response latencies in proactive interference. Participants were exposed to lists of three items, with blocks of three lists involving words from the same semantic category. Across a block's three lists proactive interference built up: with increasing number of already learned lists, recall performance declined. More interesting, with increas-

ing number of already learned lists, mean latency (τ) increased. Consistent with previous computational accounts of interference (Mensink & Raaijmakers, 1988; Raaijmakers & Shiffrin, 1981), this increase in response latency suggests that interference already acts at the sampling stage of recall: The encoding of the previously studied material reduces the relative strength of the more recently studied material and thus lowers the material's chances for being sampled.

Part-list cuing refers to the robust observation that the presentation of a subset of previously learned material as a retrieval cue can have a detrimental effect on recall of the remaining material (for reviews, see Nickerson, 1984; Roediger & Neely, 1982). A leading theory of part-list cuing is still retrieval competition that attributes the forgetting to the sampling stage of recall (Rundus, 1973). In this account, reexposure of items as cues introduces a sampling bias that reduces recall totals of noncue items and, at the same time, increases the items' response latencies. Roediger, Stellon, and Tulving (1977) analyzed recall totals and response latencies in part-list cuing. They found the typical detrimental effect of part-list cues on recall totals. In addition, they found slower mean latencies (τ) in the presence rather than the absence of part-list cues, which is consistent with the retrieval competition account and attributes the forgetting to the sampling stage of recall.

Directed forgetting has also been explained in terms of sampling. In the list method of directed forgetting, participants study two lists of items. After the presentation of List 1, they receive a cue to either forget or continue remembering this list while studying List 2. Compared with remember-cued participants, forget-cued participants typically show impaired recall of List 1 and improved recall of List 2 items (for a review, see MacLeod, 1998). This finding has often been proposed to be the result of retrieval inhibition. Geiselman et al. (1983), for instance, suggested that a cue to forget a previously learned list blocks the access routes to the episodic memory traces, making them unretrievable. The to-be-forgotten items are excluded from the task's search set, and forgetting arises without affecting the items' representation itself. However, no response latency analysis has yet been conducted to test this unbinding account of directed forgetting directly.

Sampling and Recovery in Retrieval-Induced Forgetting

Although all of the accounts of episodic forgetting mentioned above differ in detail, they all attribute the forgetting to the sampling stage of recall. Arguably, retrieval-induced forgetting might also reflect a sampling problem. Such a suggestion arises from both inhibitory and noninhibitory accounts of the phenomenon. Williams and Zacks (2001), for instance, questioned the inhibitory account of retrieval-induced forgetting and rather preferred a retrieval interference explanation. According to their explanation, the forgetting does not occur in the retrieval-practice phase but rather in the testing phase of the experiment, in which recall of the practiced items supposedly blocks recall of the unpracticed material. This explanation, which shows a strong resemblance to the retrieval competition account of part-list cuing, indicates that retrieval-induced forgetting, just like interference and part-list cuing, should be located at the sampling stage of recall.

A second reason to argue for a sampling account of retrieval-induced forgetting arises from the fact that retrieval-induced for-

getting, just like directed forgetting, has been attributed to retrieval inhibition. If, as in Geiselman et al.'s (1983) proposal, inhibition caused an unbinding process and reduced the interference potential of competing items by disconnecting the items from the common cue, then the inhibitory processes underlying retrieval-induced forgetting might also reflect an unbinding process, and retrieval practice might block the retrieval routes to interfering items to prevent them from disrupting retrieval of the practiced items. Indeed, in the episodic forgetting literature, it has repeatedly been argued that retrieval inhibition may be realized by means of such an unbinding mechanism (Postman et al., 1968).

The possible conclusion that retrieval-induced forgetting reflects a failure at the sampling stage of recall might be premature, however. Indeed, results from prior work challenge Williams and Zacks' (2001) retrieval interference account by showing that the forgetting does not only occur in category-cued recall tasks but also in word-stem completion (Anderson et al., 1994; Bäuml & Hartinger, 2002), tests of recognition memory (Hicks & Starns, 2004; Starns & Hicks, 2004), and even implicit memory tests (Perfect, Moulin, Conway, & Perry, 2002; Veling & van Knippenberg, 2004; however, see Butler, Williams, Zacks, & Maki, 2001). Because at least some of these tests include a rough control of sampling biases (Perfect et al., 2002; Veling & van Knippenberg, 2004; for a discussion, see also Bäuml & Aslan, 2004), the results suggest that retrieval-induced forgetting may not reflect a sampling problem.

Moreover, the assumption that retrieval-induced forgetting is caused by retrieval inhibition does not necessarily imply that the forgetting reflects the action of an unbinding mechanism. Although unbinding would be effective in reducing interference for the target material, the proposal is in conflict with the cue independence finding of retrieval-induced forgetting (Anderson & Bell, 2001; Anderson & Green, 2001; Anderson & Spellman, 1995; Veling & van Knippenberg, 2004). Cue independence refers to the observation that retrieval-induced forgetting does not only arise if, at test, the same cues are provided as are used during the retrieval-practice phase. Rather, the forgetting generalizes to the case in which cues are exposed that are not associated to the practiced item and, in this sense, reflect independent cues. According to cue independence, it is not damage to any particular association that is at the heart of retrieval-induced forgetting, thus challenging an unbinding interpretation of retrieval-induced forgetting (however, for failures to find cue independence, see Perfect et al., 2004; Williams & Zacks, 2001).

Indeed, retrieval-induced forgetting has also been thought of as a recovery problem. It has been argued that, in retrieval-induced forgetting, inhibition may reduce the level of activation for a given item and thus prevent the item from achieving the recovery threshold (Anderson, 2003). Supporting this view, Veling and van Knippenberg (2004) measured reaction times in recognition and implicit memory tasks, instructing the participants to react as quickly as possible to the presented stimuli. Presenting participants with the items themselves should preempt the active retrieval process, and the resulting reaction times should be indicative of the activation level of the items. In both types of tasks, reaction times for unpracticed items increased, which is consistent with the proposal that inhibition reduces the activation level of items. Because active retrieval was largely eliminated in this study, the results, however,

are silent about the possible role of sampling processes in retrieval-induced forgetting.

A recovery account of retrieval-induced forgetting would provide an explanation for the cue independence finding and the persistence of the effect in word-stem completion, item recognition, and implicit tests. However, it would contrast sharply with accounts of other forms of episodic forgetting, which typically attribute episodic forgetting to the sampling stage of recall, and thus would suggest that retrieval-induced forgetting reflects a unique form of episodic forgetting. The current study arrives at a time when the existing data do not allow a clear-cut answer on whether retrieval-induced forgetting reflects a sampling failure, a recovery failure, or both. By analyzing response latencies of practiced and unpracticed items, the experiments reported below address the issue directly.

Experiment 1

The results of two experiments are reported in which it is examined whether the detrimental effects of retrieval practice are located at the sampling or the recovery stage of recall. In Experiment 1, participants learned items from different semantic categories, with half of the items of a category belonging to one semantic subcategory and the other half to another. In a subsequent retrieval-practice phase, participants practiced retrieval on a subset of the categories by repeatedly retrieving the items of one subcategory; the items of the categories' other subcategory were not practiced (see Figure 1). As has recently been shown (Bäuml & Hartinger, 2002), retrieval practice on the items of one subcategory can cause forgetting of the other subcategory's items. At test, participants received the category plus subcategory labels of the items as cues and recalled the items that belonged to the cues. Because the practiced and unpracticed items were from different semantic subcategories, they could be tested separately. We measured recall total and response latencies of the items.

Concerning recall total, we expected to find the standard pattern of retrieval-induced forgetting, with improved recall of the practiced and impaired recall of the unpracticed items. More interestingly, we examined whether retrieval practice influenced the items' response latencies as well. If retrieval practice influenced recall total of the unpracticed items but not their response latencies, then this would be evidence that the effect of retrieval
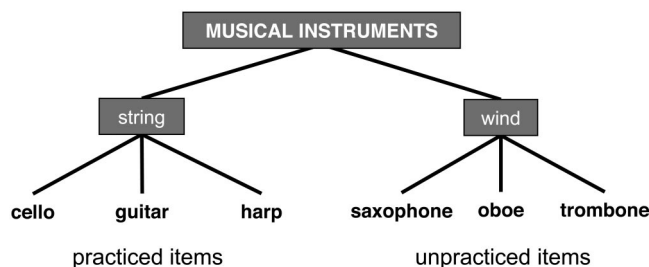

Figure 1. Example of a category as used in Experiment 1. Participants studied six items from the category, with three items belonging to one subcategory and three items belonging to another. In a subsequent retrieval-practice phase, only the items of one subcategory were practiced, leading to retrieval inhibition of the unpracticed items and retrieval-induced forgetting on a later recall test.

practice occurs at the recovery stage of recall. This result would be consistent with the view that retrieval practice reduces the items' activation level and thus prevents them from exceeding the recovery threshold.

Indeed, as illustrated in the top panel of Figure 2, if a reduction in activation level mediated the forgetting, then, with and without inhibition, the search set connected to the cue wind would consist of the three items saxophone, oboe, and trombone, which would make the probability for an item to be sampled one third in both cases. Because this value, together with the (normal) sampling rate, determines mean latency, mean latency would not be affected. Even so, fewer items would be recalled with than without inhibition, because items that are reduced in their activation level are less likely to exceed the recovery threshold.

If retrieval practice affected recall total and response latencies of the unpracticed items, then this would be evidence that the effect of retrieval practice occurs at the sampling stage of recall. As illustrated in the middle panel of Figure 2, if response latencies were slower for the unpracticed items, then it would be indicated that retrieval practice reduces the speed of processing of unpracticed items, thus mimicking effects of negative priming or, following the noninhibitory account of the phenomenon, introduces a sampling bias for the unpracticed items. Both mechanisms would not affect the size of the search set—that is, the number of items connected to the cue wind—and would not change the three items' activation level.

If retrieval practice made response latencies for the unpracticed items faster, then this would be evidence for a reduction in the size of the mental search set and the action of an unbinding mechanism. As illustrated in the bottom panel of Figure 2, if unbinding medi-

ated the forgetting, then the search set connected to the cue wind would no longer consist of the three items saxophone, oboe, and trombone but, for instance, would consist only of the two items saxophone and trombone. This would increase the probability for an item to be sampled to one half and, as a result, faster mean latencies would arise with than without inhibition. The results of the experiment thus speak to the issue of how inhibition is realized in retrieval-induced forgetting.

Method

Participants. A total of 24 students at the University of Regensburg (Regensburg, Germany) participated in the experiment. They were tested individually.

Material. An item list was constructed consisting of six experimental and two filler categories. Each category included three items from each of two semantic subcategories. The category tree, for instance, contained three exemplars belonging to the subcategory deciduous tree and three exemplars belonging to the subcategory conifer; the category professions contained three exemplars belonging to the subcategory academics and three exemplars belonging to the subcategory workmen. The items were drawn from several published norms (Battig & Montague, 1969; Mannhaupt, 1983; Scheithe & Bäuml, 1995). Items from rank order 1–4 were excluded to prevent guessing. Strong associations between items and strong similarities between categories were avoided. The two initial letters of each item were unique with respect to that item's category.

Design. The experiment was conducted in three main phases separated by distractor tasks. In Phase 1, participants studied the item list. After a subsequent distractor task, a retrieval-practice phase followed (Phase 2) in which, from four of the six experimental categories, the items of one subcategory were practiced; neither the items from the categories' other subcategory were practiced nor were the items from the two remaining
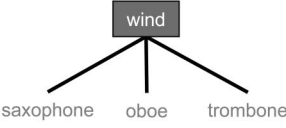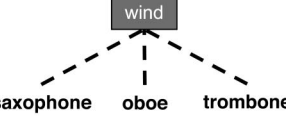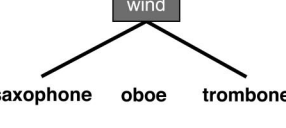

Inhibitory mechanisms	Search set	x	Sampling rate	=	Mean latency (τ)	Activation level	Type of failure
Reduction in activation level	normal		normal		normal	reduced	recovery
Retrieval slowing	normal		slow		slow	normal	sampling
Unbinding	small		normal		fast	normal	sampling

Figure 2. Effects of different putative inhibitory mechanisms—reduction in activation level, retrieval slowing, and unbinding—on mean latency (tau [τ]) and activation level of the unpracticed items, together with the recall stage at which the failure according to sampling/recovery models occurs.

experimental categories and the two filler categories. After another distractor task, participants completed a final category-cued-recall test in which the items studied in Phase 1 of the experiment had to be recalled (Phase 3). Both the category and the subcategory labels were provided as retrieval cues. Using such a testing procedure, participants recalled the practiced and unpracticed items separately. This design created three types of items. From two experimental categories, no items were practiced; these items define the baseline items in the experiment—not retrieval practiced (Nrp) items. From four experimental categories, the items of one subcategory were practiced; the items from the categories' practiced subcategories define the retrieval practice plus (Rp+) items; the items from the categories' unpracticed subcategories define the retrieval practice minus (Rp−) items.

Procedure. In the study phase, a list consisting of 36 (6 × 6) experimental and 12 (2 × 6) filler items was presented to the participants. Each item, together with its category and subcategory label, was displayed on a computer screen for 4 s (e.g., profession academic lawyer). The item was presented underneath the subcategory label, which in turn was presented underneath the category label. The experimental items were presented in blocked random order. That is, a random sequence of six blocks with six items each was presented to the participants. Each block consisted of one randomly selected exemplar from each of the six categories. The order of the categories within a block was also random, with the only restriction that a block's last item never belonged to the same category as the next block's first item. At the beginning and end of the list, two items from each of the two filler categories were presented. The remaining four filler items were randomly interspersed in the list.

After 60 s of a number-ordering task, a retrieval-practice phase followed, in which the items from four subcategories (Rp+ items) had to be retrieved. The retrieval of the items was controlled by presenting the category label plus the two (unique) initial letters of the item as retrieval cues. The subcategory labels were not provided. This was done to maximize interference from the other subcategory's items and thus give rise to inhibition (Bäuml & Hartinger, 2002). The Rp+ items were retrieved in two consecutive blocks with no break between the blocks. Each Rp+ item was retrieved once within each block. The order of the items within a block was random. Each retrieval cue was presented for 7 s on the computer screen. During this time, the participants answered with the target item that was recorded by the experimenter. Subsequent to the retrieval-practice procedure, an additional 2-min distractor task followed in which some math problems had to be solved.

In the test phase, participants were asked to recall the items studied in Phase 1 of the experiment. As retrieval cues, both the category and the subcategory labels of the items were provided. The retrieval cues were displayed on the computer screen for 20 s. During this time, the participants said aloud the items belonging to the retrieval cues. The tested-first subcategory was always a filler category to make participants familiar with the task. The remaining filler categories were tested at the end of the recall phase. Testing order of experimental subcategories followed a blocked randomization, so that in each block, each type of item (Rp+, Rp−, Nrp) was tested once. The participants' answers were recorded by a computer program in a pulse code modulation—waveform format with a sampling rate of 44.1 kHz and a 16-bit resolution. Latencies were assessed by means of the computer program Cool Edit 2000 (Version 4.1, Syntrillium Software Corporation, Phoenix, Arizona), whereby the voice onset of each recalled item was manually located in the spectrogram. On average, errors (repetitions and false alarms) occurred in about 5 per 100 correct responses for the three item types and were excluded from the analyses.

Measure of latency. For each of the three item types (Rp+, Rp−, Nrp), two measures of response latency are presented: first-response latency and subsequent-response latency. First-response latency measures the time until the first response occurred within a subcategory. The time prior to the first response includes establishing the search set and starting the memory search and is not essential for the retrieval process itself. Therefore,

first-response latencies and subsequent-response latencies are often analyzed separately. Subsequent-response latency is measured by determining, for each subsequently recalled item, the time elapsed since the first response rather than the time elapsed since the beginning of the recall period. In this way, a more accurate portrayal of the time course of retrieval is obtained (for a detailed discussion on this point, see Rohrer et al., 1995).

To analyze retrieval dynamics, we fitted exponential functions to the subsequent-response latency functions of each item type (see Introduction). The functions are described by two parameters—N representing asymptotic recall, and τ representing the mean latency of those N items—that were derived from fitting the exponential to the data. The best-fitting exponentials were determined by least-square minimization. Using the asymptotic standard error for each parameter, we performed pairwise comparisons of parameter values by a t test. For these t tests, the asymptotic standard error of each parameter value provided a measure of the variability of each parameter, and the degrees of freedom for each of the two curve fits, summed together, provided the number of degrees of freedom (for details, see Rohrer et al., 1995).[2]

Results

Recall totals. In the retrieval-practice phase, 94.8% of the items were correctly completed, indicating that in the overwhelming majority of cases, retrieval of the Rp+ items was successful.

In the test phase, on average, 89.2% of the Rp+ items, 58.7% of the Rp− items, and 69.1% of the Nrp items were recalled. The difference of 20.1% between recall of the Rp+ items and recall of the Nrp items was reliable, $F(1, 23) = 52.0$, $p < .001$, demonstrating the expected positive effect of retrieval practice on recall of the Rp+ items. The difference of 10.4% between recall of the Rp− items and recall of the Nrp items was also reliable, $F(1, 23) = 9.1$, $p < .01$, showing the expected negative effect of retrieval practice on recall of the Rp− items. Thus, observed recall totals revealed the standard pattern of retrieval-induced forgetting.

Response latencies. Table 1 shows the first-response latencies of the three item types. Mean first-response latencies were 2.5 s for the Rp+ items, 3.0 s for the Rp−items, and 3.1 s for the Nrp items. The difference of 0.6 s between the Rp+ and Nrp items was significant, $F(1, 23) = 10.1$, $p < .01$, the difference of 0.1 s between the Rp− and Nrp items was not, $F(1, 23) < 1.0$.

Subsequent-response latencies were grouped into 2-s bins and plotted as a function of time (see Figure 3). Specifically, for each of the three item types, each data point represents the average percentage of items that were produced in that 2-s bin. Figure 3 also shows the best-fitting, two-parameter exponential for each of the three item types. The exponential provides a good description of the three data sets and, as reported in Table 1, accounts for a large portion of the variance. The parameter estimate of asymptotic percentage (N) revealed values of 86.3% for the Rp+ items, 43.2% for the Rp− items, and 54.3% for the Nrp items. Because N is based on subsequent responses only, whereas recall total includes first responses as well, we computed corrected totals, in which

---

[2] Rather than analyzing subsequent latencies in terms of mean latencies, we could have analyzed them in terms of interresponse times, that is, the time between successive recalls. However, there are at least two good reasons to use mean latencies for analysis. First, mean latency is what the exponential τ parameter captures. Second, mean latency is theoretically interesting, because τ is supposed to vary linearly with the size of the search set (for details and further arguments for the use of mean latency, see Wixted & Rohrer, 1994).

Table 1
Recall Performance (Percentage Recalled) and Response Latencies (in Seconds) for Experiment 1

Item type	% recalled		First-response latency		Subsequent-response latency (τ)		
	M	SE	M	SE	M	SE	VAF
Rp+	89.24	2.10	2.48	0.10	2.17	0.07	0.99
Rp−	58.68	3.38	2.95	0.15	2.52	0.11	0.99
Nrp	69.10	2.16	3.05	0.18	2.68	0.19	0.99

Note. Subsequent response latency excludes time until first response and is estimated by an exponential (see text). VAF = variance accounted for by the exponential; Rp+ = retrieval practice plus; Rp− = retrieval practice minus; Nrp = not retrieval practiced.

only the subsequent responses were included. Corrected totals—83.9% for the Rp+ items, 42.2% for the Rp− items, and 53.7% for the Nrp items—were very similar to the estimates of *N*, indicating that recall was close to asymptote in the current experiment, a feature well reflected in Figure 3.

Of primary interest in the current study are the parameter estimates of mean subsequent-response latency (τ). These values were 2.2 for Rp+ items, 2.5 for Rp− items, and 2.7 for Nrp items, as reported in Table 1. The difference of 0.5 s between the Rp+ and Nrp items was significant, $t(8 + 8) = t(16) = 2.6$, $p < .05$, whereas the difference of 0.2 s between Rp− and Nrp items was not, $t(16) < 1.0$. Thus, retrieval practice influenced the response latencies of Rp+ items—first and subsequent responses—but did not affect the latencies of Rp− items.

Further data and analyses. A rationale behind Experiment 1 was that the presentation of a subcategory cue at test activates only items from this subcategory and not from the category's other subcategory, thus assuming that a category's practiced and unpracticed items belonged to different search sets. If true, then cuing

Figure 3. Retrieval-induced forgetting with semantic categories (Experiment 1): Recall performance—percentage recalled—for each 2-s bin for the retrieval practice plus (Rp+), retrieval practice minus (Rp−), and not retrieval practiced (Nrp) items together with the best-fitting exponentials. Latency is measured from the first response.

with subcategory labels should create much smaller search sets, and faster response latencies (τ), than cuing with the superordinate category labels (Rohrer, 1996; Wixted & Rohrer, 1994). We checked for this property in a control experiment, in which we repeated Experiment 1 with two changes. The one change was that, between learning and test, there was no retrieval practice but rather an extended distractor phase. The other change was that, at test, one half of the participants received the category cues only and were asked to recall all of the categories' items; the other half received the category plus subcategory cues and were asked to recall the items that belonged to the two cues. Participants had 40 s to recall a category's items and 20 s to recall a subcategory's items. Sixteen participants took part in the experiment.

As indicated in Table 2, the two groups did not differ in recall totals, $F(1, 14) < 1.0$, and they did not differ in first-response latencies, $F(1, 14) = 1.4$, $p > .20$. Figure 4 shows subsequent-response latencies for the two groups. Obviously, responses were much faster in the subcategory group than the category group. Consistently, fitting the two-parameter exponential to the data of the two groups yielded τ values of 2.1 for the subcategory group and 5.6 for the category group, $t(8 + 8) = t(16) = 5.7$, $p < .001$. This difference in latencies indicates that the cuing with subcategory labels in Experiment 1 really led to a separation of the two subcategories' items and thus to different search sets for the practiced (Rp+) and unpracticed (Rp−) items.

A very compelling intuition one might have with the results from this control experiment is that the difference in mean latency (τ) was not due to a difference in the search set size but rather was caused by the difference in the absolute number of items recalled. Indeed, whereas recall percentages were identical across the two testing conditions, the absolute number of items recalled was not. Two lines of evidence indicate that the difference in latencies was caused by the difference in search set size. One line of evidence comes from studies showing that a change in recall total is not necessarily accompanied by a change in mean latencies. Rohrer and Wixted (1994), for instance, increased recall total through an increase in items' study time without affecting the items' mean latencies. Similarly, Experiment 1 showed that a decrease in unpracticed items' recall total caused by retrieval practice of related items is not accompanied by a decrease in the items' response latencies.
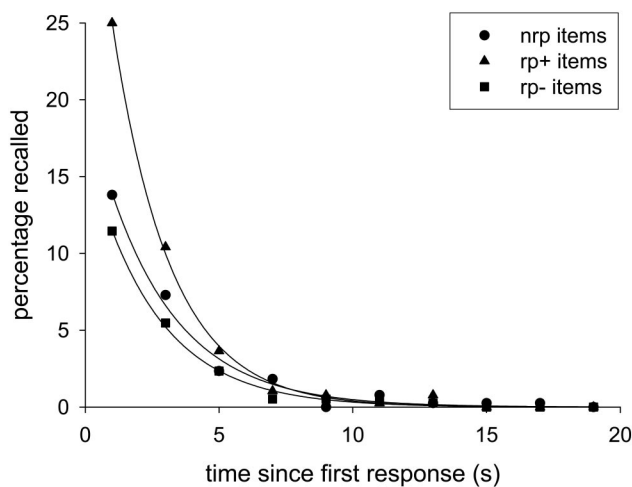
Table 2
Recall Performance (Percentage Recalled) and Response Latencies (in Seconds) for the Control Experiment

Type of cue	% recalled		First-response latency		Subsequent-response latency (τ)		
	M	SE	M	SE	M	SE	VAF
Category	60.42	3.74	3.21	0.24	5.63	0.56	0.94
Subcategory	57.29	2.92	2.73	0.32	2.14	0.24	0.97

Note. Subsequent response latency excludes time until first response and is estimated by an exponential (see text). VAF = variance accounted for by the exponential.

A second line of evidence rests on the prediction of the random search model that mean latency (τ) depends linearly on search set size (McGill, 1963). Indeed, as for instance shown by Rohrer and Wixted (1994), mean latencies double if the number of studied items—and thus the number of to-be-recalled items—doubles. The absence of the subcategory labels in the current experiment also doubled the number of to-be-recalled items. If this doubling was accompanied by a doubling of the search set, then mean latencies in the category condition should have been roughly twice as large as mean latencies in the subcategory condition. This is exactly what the data show, consistent with the view that the introduction of the subcategory labels led to separate search sets for the practiced (Rp+) and unpracticed (Rp−) items.[3]

Discussion

Retrieval practice resulted in higher recall totals for the practiced items and lower recall totals for the unpracticed items, thus replicating the standard pattern of retrieval-induced forgetting. Although retrieval practice lowered recall total of unpracticed items, it did not affect their response latencies. This held true for

first-response latencies and subsequent-response latencies. Finding no difference in subsequent-response latencies between unpracticed and control items indicates that retrieval inhibition should not have acted at the sampling stage of recall in this experiment but rather at the recovery stage. This result is consistent with the view that inhibition reduces the activation level of unpracticed items, so that they no longer exceed the recovery threshold.

Indeed, referring back to the top panel of Figure 2 and the three items belonging to the wind search subset, a reduction-in-activation-level mechanism does not affect the size of the wind search subset and, therefore, does also not affect the probability of the inhibited items to be sampled. Because the mechanism does also not influence sampling rate, the whole sampling process remains intact and mean latency (τ) does not change. Because the activation level of the single items is reduced, however, the recovery process is affected. Fewer items are recovered into consciousness, and, as a result, recall totals are reduced.

As opposed to the unpracticed items, retrieval practice affected recall total and response latencies of practiced items. Practiced items showed slightly faster first responses and slightly faster subsequent responses than control items. Because the time prior to the first response includes establishing the search set and starting the memory search, this result indicates that retrieval practice led to a faster start of the memory search in practiced items than in control items. Retrieval practice also influenced the retrieval process itself, presumably by increasing the speed of processing in retrieval of these items. Thus, at least part of the positive effect of retrieval practice should have occurred at the sampling stage of recall.
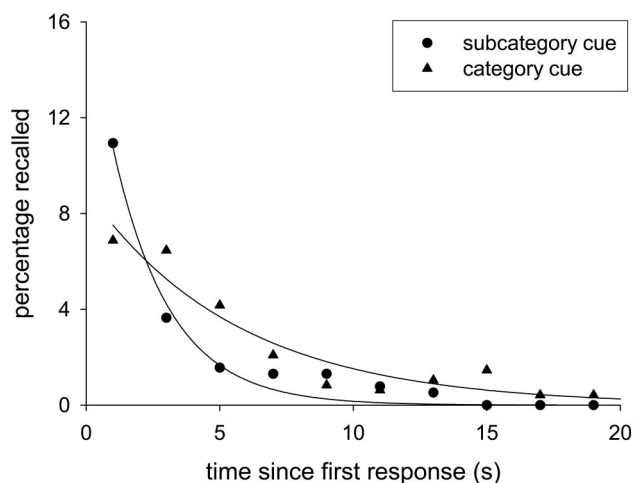

Figure 4. Effect of cuing with the subcategory label (control experiment): Recall performance—percentage recalled—for each 2-s bin for the category group and the subcategory group together with the best-fitting exponentials. Data for the category group are shown only up to the 20th second. Latency is measured from the first response.

---

[3] Prior work has shown that recall of a category can be affected by the other categories on the study list (e.g., Tulving & Pearlstone, 1966), suggesting that items from other categories may have been part of the search set of the Rp+ and Rp− items in the current experiments. It appears unlikely that such effects influenced the current results in a substantial way. First, effects of other categories on the study list have typically been found to be small, if existent at all (e.g., Tulving & Psotka, 1971). Second, the elimination of the subcategory label in the control experiment (more than) doubled mean latencies, which indicates that it roughly doubled search set size. If a major number of items from other categories had been part of the search set, then the elimination of the subcategory label should not have doubled search set size.

Experiment 2

Results from both interference and part-list cuing studies showed a reduction in recall total accompanied by longer response latencies. The results of Experiment 1 show a different pattern: Retrieval practice reduced recall total of the unpracticed items, but it did not affect the items' response latencies. Following sampling/recovery models of recall, these findings indicate that only some forms of episodic forgetting, like interference and part-list cuing, may be located at the sampling stage of recall, whereas others, like retrieval-induced forgetting, may be located at the recovery stage. The goal of Experiment 2 was to replicate the result of Experiment 1 with different material and a combination of spatial and semantic cues rather than purely semantic cues.

In Experiment 2, participants studied two sets of items, one set presented on the left side of a computer screen, the other set presented on the right side. Each set consisted of items from different semantic categories. In a subsequent retrieval-practice phase, participants practiced retrieval on the items from half of the categories of one (left) set; the items from the other (right) set were not practiced. Following previous output interference studies, in which it was shown that retrieval-induced forgetting can cross category boundaries (Roediger & Schmidt, 1980; Smith, 1971; see also Anderson & Spellman, 1995), we expected that the retrieval practice caused forgetting of the unpracticed items from the same (left) set. Following other previous research (Basden & Basden, 1995), we did not expect that retrieval practice caused forgetting of the items from the other (right) set. At test, participants received the spatial cue (e.g., left) together with the category cue (e.g., furniture) of the items and were asked to recall the items that belonged to the cues. Because a set's practiced and unpracticed items were from different semantic categories, they could again be tested separately. Like in Experiment 1, we measured recall total and response latencies of the items.

Method

Participants. Twenty-four students at the University of Regensburg participated in the experiment. They were tested individually.

Material. An item list was constructed consisting of two sets of seven semantic categories each. Six of the categories were experimental categories, the seventh category was a filler category. For each category, four items were selected from published norms (Battig & Montague, 1969; Mannhaupt, 1983; Scheithe & Bäuml, 1995). The items were chosen to have a rank order between 5 and 20, according to these norms. The two initial letters of each item were unique with respect to the item's set.

Design. The design was similar to the design of Experiment 1. The experiment was conducted in three main phases separated by distractor tasks: a study phase, a retrieval-practice phase, and a test phase. The main difference across experiments was in the construction and presentation of the item list. Whereas in Experiment 1, each item was associated to one category and one subcategory cue, in Experiment 2, each item was associated to one spatial and one category cue. Spatial cues were introduced by presenting one set of items in the middle of the left half and the other set in the middle of the right half of the computer screen, and instructing participants to think of the items as a left and right column's items (Basden & Basden, 1995). In the retrieval-practice phase, the items from half of the categories of one set were practiced; the items from the other set were not practiced. In the test phase, participants were asked to remember the items learned in the study phase, given the spatial and the category labels as

retrieval cues. This design again created three types of items. The items belonging to the practiced categories are the Rp+ items. The items belonging to the same set's unpracticed categories define the Rp− items. The items from the other set's unpracticed categories define the Nrp items, which serve as the baseline items in the experiment.

Procedure. In the study phase, the item list, which consisted of the two sets of 28 (7 × 4) items each, was presented to the participants. To facilitate separation of the two sets of items, we printed the items in different colors, dependent on which side of the screen they were presented. To enhance interference between a set's items, we showed the items without their category labels. Throughout the whole presentation phase, the labels left and right were displayed above the items' screen locations. Items were presented in succession, with each item shown for 5 s. The two sets' items were presented alternately. As in Experiment 1, item order within a set followed a blocked randomization. The assignment of the two sets of items to the two spatial locations was balanced across participants.

After 60 s of a number-ordering task, a retrieval-practice phase followed, in which, from one of the two sets of items, half of the categories (Rp+) had to be retrieved. The retrieval of the items was controlled by presenting the spatial cue (left or right) plus the items' two unique initial letters as retrieval cues. The items' category labels were not provided. Analogous to Experiment 1, this was done to enhance interference from the set's other categories and thus give rise to inhibition. In all other aspects, the retrieval-practice phase was identical to the retrieval-practice phase of Experiment 1. Participants had 7 s to answer with the correct item, and items were practiced in two consecutive blocks. Subsequent to the retrieval-practice procedure, a 2-min distractor task followed in which some math problems had to be solved.

In the test phase, participants were asked to remember the items learned in the first phase of the experiment. As retrieval cues, the spatial label plus the category label of the items were displayed on the computer screen for 20 s (e.g., left furniture), and participants had to say aloud the cued category's items during this time. The tested-first and tested-second categories were always the two filler categories that were not included in the data analysis. This was done to familiarize participants with the testing procedure. Becoming familiar with the testing procedure was important because, at this point, the presentation of category labels might have been unexpected for some participants. Testing order of categories followed a blocked randomization so that in each block, one Rp+, one Rp−, and two Nrp categories were tested. The participants' answers were recorded and analyzed in the same way as in Experiment 1.

Results

Recall totals. In the retrieval-practice phase, 85.4% of the items were correctly completed, indicating that in the majority of cases, retrieval of the Rp+ items was successful.

In the test phase, on average, 76.0% of the Rp+ items, 47.6% of the Rp− items, and 58.7% of the Nrp items were recalled. The difference of 17.3% between recall of the Rp+ items and recall of the Nrp items was reliable, $F(1, 23) = 20.1$, $p < .001$, as was the difference of 11.1% between recall of the Rp− items and recall of the Nrp items, $F(1, 23) = 12.5$, $p < .01$. Observed recall totals thus revealed the standard pattern of retrieval-induced forgetting.

Response latencies. Mean first-response latencies were 2.9 s for the Rp+ items, 3.7 s for the Rp− items, and 3.2 s for the Nrp items. The difference of 0.3 s between the Rp+ and Nrp items was marginally significant, $F(1, 23) = 3.1$, $p < .10$, as was the difference of 0.5 s between the Rp− and Nrp items, $F(1, 23) = 2.9$, $p = .10$. The difference of 0.8 s between the Rp+ and Rp− items was significant, $F(1, 23) = 2.7$, $p < .05$.

Subsequent-response latencies were again grouped into 2-s bins and plotted as a function of time (see Figure 5). Thus, for each of the three item types, each data point represents the average percentage of items that were produced in that 2-s bin. Figure 5 also shows the three best-fitting exponentials. For each item type, the exponential provides a good description of the data set and, as shown in Table 3, accounts for a large portion of the variance for each item type. The parameter estimate of asymptotic percentage (N) revealed values of 71.9% for the Rp+ items, 33.8% for the Rp− items, and 44.8% for the Nrp items. These values are again smaller than those of observed totals, which include first responses as well. Computing corrected values, in which only subsequent responses are included, led to recall totals of 68.5% for the Rp+ items, 34.3% for the Rp− items, and 45.8% for the Nrp items. These values are close to the estimates of N, indicating that recall was close to asymptote. This pattern is also reflected in Figure 5.

Of primary interest, parameter estimates of mean subsequent-response latency (τ) were 3.1 for Rp+ items, 2.6 for Rp− items, and 2.5 for Nrp items, as reported in Table 3. The difference of 0.6 s between the Rp+ and Nrp items was not significant, $t(8 + 8) = t(16) = 1.7$, $p > .10$, and neither was the difference of 0.1 s between Rp− and Nrp items, $t(16) < 1.0$. Retrieval practice thus had no reliable influence on subsequent-response latencies. This holds, although the size—not the direction—of the difference between Rp+ and Nrp items was well comparable with the significant effect found in Experiment 1.[4]

Discussion

Experiment 2 showed largely the same negative effects of retrieval practice on unpracticed items as were found in Experiment 1. Although retrieval practice reduced recall total, it did not affect the items' subsequent latencies. Thus, retrieval inhibition should not have acted at the sampling stage but rather at the recovery stage of recall, consistent with the view that inhibition

Figure 5. Retrieval-induced forgetting with spatial categories (Experiment 2): Recall performance—percentage recalled—for each 2-s bin for the retrieval practice plus (Rp+), retrieval practice minus (Rp−), and not retrieval practiced (Nrp) items together with the best-fitting exponentials. Latency is measured from the first response.

reduces the activation level of items. Besides, there was a marginally significant effect of retrieval practice on unpracticed items' first responses, indicating that it took a bit longer to establish the search set and start the memory search for these items. The effect, however, was small.

Again, retrieval practice improved later recall of the practiced items. This time, the recall improvement was not accompanied by a significant effect on the items' mean latencies, suggesting that retrieval practice did not affect the retrieval process of practiced items itself. Retrieval practice thus should have increased the activation level of the practiced items, very similar to how it reduced the activation level of the unpracticed items. There was again a small and marginally significant effect of retrieval practice on practiced items' first responses. The effect indicates that starting the memory search was slightly faster for practiced items than for unpracticed items.

General Discussion

Retrieval-Induced Forgetting as a Recovery Failure

In two experiments, we examined retrieval-induced forgetting by analyzing unpracticed items' recall totals and response latencies. We replicated prior work by showing that retrieval practice on a subset of previously learned items reduces recall total of the unpracticed items (Anderson, 2003). We extended prior work by demonstrating that the reduction in recall total is not accompanied by an effect on the items' subsequent-response latencies and is accompanied by only a small effect, if at all, on items' first-response latencies. Retrieval practice thus had a different influence on observed total and response latencies.

First-response latencies provide information on how long it takes for a certain retrieval cue to delimit the mental search set and start the memory search, whereas subsequent-response latencies address the retrieval process itself. The fact that, in Experiment 1, unpracticed items did not differ significantly in their first-response latencies from control items and, in Experiment 2, showed only a tendency for slower first-response latencies indicates that retrieval practice had at best a minor influence on the initiation of the memory search for the inhibited items.

The result that retrieval practice did not affect unpracticed items' subsequent responses indicates that retrieval practice did not influence the sampling of inhibited items. Influences on sampling should lead to either a decrease or an increase in response latencies. A decrease, for instance, would occur if inhibition was due to an unbinding mechanism and was caused by the disconnection of interfering items from the common cue (Geiselman et al., 1983; Postman et al., 1968; see Figure 2, bottom panel). In this
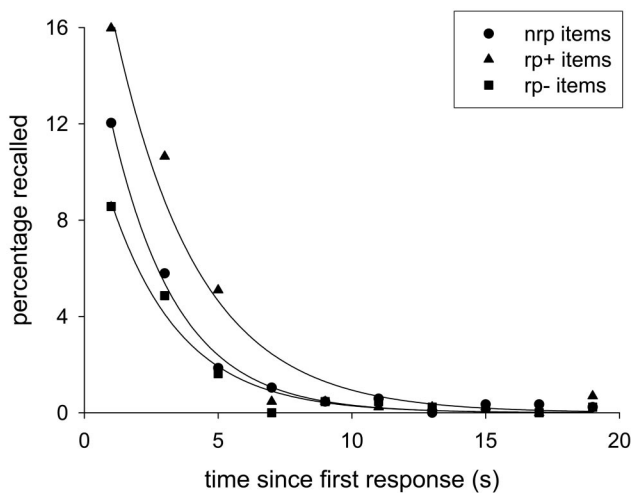
---

[4] We estimated parameters also for cumulative latency functions. In Experiment 1, estimates of τ were 2.2 for Rp+ items, 2.5 for Rp− items, and 2.7 for Nrp items; in the control experiment, estimates of τ were 5.3 for the category group and 3.0 for the subcategory group; in Experiment 3, estimates of τ were 2.7 for Rp+ items, 2.5 for Rp− items, and 2.6 for Nrp items. Although the parameters thus were not exactly identical to those estimated with the noncumulative functions, the pattern of results and the statistical conclusions remained the same. The same holds true when weighted least-square methods are used rather than unweighted methods in the analyses.

Table 3
Recall Performance (Percentage Recalled) and Response Latencies (in Seconds) for Experiment 2

Item type	% recalled		First-response latency		Subsequent-response latency (τ)		VAF
	M	SE	M	SE	M	SE	
Rp+	76.04	3.86	2.87	0.19	3.13	0.37	0.96
Rp−	47.57	3.63	3.66	0.33	2.63	0.27	0.97
Nrp	58.68	2.10	3.17	0.14	2.47	0.13	0.99

Note. Subsequent response latency excludes time until first response and is estimated by an exponential (see text). VAF = variance accounted for by the exponential; Rp+ = retrieval practice plus; Rp− = retrieval practice minus; Nrp = not retrieval practiced.

case, retrieval practice would reduce the effective search set and thus reduce mean latencies. An increase in latencies would occur if inhibition reduced the sampling speed of inhibited items and thus slowed item retrieval (Conway et al., 1999; Tipper & Cranston, 1985; see Figure 2, middle panel).

The current finding of a null effect on latencies disagrees with these predictions and indicates that the forgetting in the current study should not have been located at the sampling stage of recall. The results of the two experiments are theoretically consistent with the view that retrieval-induced forgetting occurs at the recovery stage of recall. According to this view, inhibition reduces the level of activation for inhibited items and thus prevents the items from achieving the recovery threshold (Anderson, 2003; Veling & van Knippenberg, 2004; see Figure 2, top panel). Inhibition, therefore, should reduce recall totals but should not affect the items' response latencies. Indeed, this is exactly what happened in the two experiments.

The current results can distinguish between different putative inhibitory mechanisms. However, they also challenge noninhibitory accounts of the phenomenon. For instance, it has been suggested that retrieval-induced forgetting is not caused by retrieval inhibition in the intermediate phase of the experiment but rather occurs in the testing phase in which recall of the practiced items blocks recall of the unpracticed material (Williams & Zacks, 2001). Recent work used electrophysiological measures of brain activity to search for retrieval-specific effects operating in the practice phase (Johansson, Aslan, Bäuml, & Mecklinger, 2005). Evidence for retrieval-specific effects was found. Complementing these results, the current data show no evidence for blocking and accompanied sampling biases in the testing phase, a pattern that, as a whole, rejects a retrieval-interference explanation of retrieval-induced forgetting.

Relation to Previous Output Interference Studies

Retrieval-induced forgetting refers to the observation that the recall of learned material can be impaired through preceding retrieval of other learned material. Evidence for such retrieval-induced forgetting was first observed in studies on output interference. In these studies, it was examined how the recall of learned items varies as a function of the items' serial position in the testing sequence. The general result of these studies was that an item's recall probability declines with its testing position (Roediger & Schmidt, 1980; Smith, 1971; Tulving & Arbuckle, 1966), thus

indicating that the act of recall itself induces the forgetting (Roediger, 1978). In several studies, it has been shown that retrieval-induced forgetting and output interference show important parallels (Bäuml, 1998; Bäuml & Hartinger, 2002).

In a recent study, we examined recall totals and response latencies in output interference (Zellner & Bäuml, 2004). Participants studied items from two target and six nontarget categories. From each category, six items were presented. Each item was shown for 3 s together with its category name. At test, the category labels were provided as retrieval cues, and, for each category, participants were asked to remember the items that belonged to the cue. One of the two target categories was tested first, the other target category was tested last. The items from the nontarget categories were tested in between, assuming that their retrieval caused inhibition of the tested-last second target category. Participants had 30 s to recall a category's items. Seventeen participants took part in the experiment.

Recall totals were significantly higher for the tested-first than the tested-last target category, thus replicating the standard finding of output interference. Regarding response latencies, there was no difference between the two categories. Neither first-response latencies nor subsequent-response latencies (τ) differed reliably between target categories. This finding is consistent with the view that output interference reflects a failure at the recovery stage of recall. In particular, it provides further evidence for the view that retrieval-induced forgetting and output interference are mediated by similar mechanisms (Bäuml & Hartinger, 2002).

Relation to Other Previous Forgetting Studies

The suggested recovery account of retrieval-induced forgetting contrasts sharply with the sampling accounts of many other forms of episodic forgetting, like interference (Raaijmakers & Shiffrin, 1981), part-list cuing (Basden & Basden, 1995; Rundus, 1973), or directed forgetting (Geiselman et al., 1983). The results from previous response latency analyses support sampling accounts of interference and part-list cuing by reporting a pattern of recall impairment accompanied by substantial increases in response latencies (Roediger et al., 1977; Wixted & Rohrer, 1994). Because, by contrast, retrieval-induced forgetting does not appear to affect response latencies, it might in fact be concluded that retrieval-induced forgetting represents a unique form of episodic forgetting. Such a conclusion appears premature, however.

Indeed, the results from recent work suggest that not only retrieval-induced forgetting but also part-list cuing may be caused by inhibition. The suggestion rests partly on results challenging the retrieval competition account of the phenomenon (Bäuml & Aslan, 2004), and partly on the results of a series of experiments in which the detrimental effects of part-list cuing and retrieval practice were compared directly, and no difference emerged (Bäuml & Kuhbandner, 2003; Zellner & Bäuml, 2005). Following the proposal that part-list cuing reflects instructed retrieval inhibition (Bäuml & Aslan, 2004), the instruction to use the cue items at test may cause covert retrieval of these items at test and thus create a sampling bias for the noncue items—which increases latencies—and additionally reduce the activation level of these items—which reduces recall totals (for a similar suggestion, see Wixted & Rohrer, 1994, p. 99). If true, then part-list cuing would reflect both a sampling and a recovery problem.

Moreover, although inhibitory accounts of directed forgetting propose that the forgetting is located at the sampling stage of recall (Geiselman et al., 1983), a recent noninhibitory account of the phenomenon (Sahakyan & Kelley, 2002) may suggest otherwise. Sahakyan and Kelley (2002) provided evidence that the benefits and costs of directed forgetting result from an internal context change that occurs between the presentation of the two lists in response to the forget instruction. Following computational models of episodic forgetting (Mensink & Raaijmakers, 1988), a context change weakens the link between each item and the retrieval context, which should lead to a reduction in recall total but not affect response latency and create a recovery rather than a sampling problem. However, no response latency analysis has yet been conducted to address the question directly. All of these results suggest that recovery problems may not only arise in retrieval-induced forgetting but in other forms of episodic forgetting as well. At least in this sense, retrieval-induced forgetting may not reflect a unique form of episodic forgetting.

Retrieval-Induced Facilitation

In both Experiment 1 and Experiment 2, we showed that retrieval practice improves recall totals of the practiced items, thus replicating results from prior work (Anderson et al., 1994; Hogan & Kintsch, 1971). We extended prior work by showing that retrieval practice can also affect practiced items' response latencies. Although the latency findings are a bit mixed across the two experiments, they suggest that the latencies of practiced and unpracticed items differ at least with respect to first-response latencies.

In Experiment 1, we found significantly shorter first-response latencies for practiced items than for control items. The same pattern arose in Experiment 2, although the effect was only marginally significant. Together, the results suggest that the initiation of the search set of the practiced items was somewhat faster than the initiation of the search set of the control items. A simple reason for the difference in first-response latencies may lie in the different effective retention intervals for the two types of items. In fact, although both types of items were part of the study phase, only the practiced items were part of the intermediate retrieval-practice phase. The effective retention interval thus was shorter for the practiced items, and this shorter retention interval may have facilitated the initiation of the search set.

Regarding subsequent-response latencies, we found significantly shorter subsequent-response latencies for practiced items than for control items in Experiment 1—the difference was on the order of 0.5 s—and nonsignificantly longer subsequent-response latencies for practiced items than for control items in Experiment 2—the difference was on the order of 0.6 s. The absence of a consistent pattern across the two experiments may suggest that there is no robust effect of retrieval practice on practiced items' subsequent latencies. However, even when focusing on the significant difference of Experiment 1, one should take into account that the size of the effect was small. As illustrated in the control study of Experiment 1 (see Table 2) and many previous studies by Wixted, Rohrer, and colleagues (Rohrer, 1996; Rohrer et al., 1995; Rohrer & Wixted, 1994; Wixted et al., 1997; Wixted & Rohrer, 1994), response latencies can vary drastically across experimental conditions, which they definitely do not in Experiment 1 or Experiment 2 (for a discussion on this point, see Rohrer, 1996). In this sense, the findings suggest that retrieval practice has at best a minor effect on the retrieval process of practiced items itself.

This suggestion is theoretically consistent with the view that retrieval practice (primarily) enhances the activation level of practiced items, thus making it easier for these items to exceed the recovery threshold and increase recall total. To the extent that this view holds true, the positive effect of retrieval practice would mimic the effect of an increase in study time or number of study trials (Rohrer, 1996; Wixted et al., 1997). Moreover, it would be indicated that both the positive and the negative effects of retrieval practice are caused through modulations in items' activation levels and thus are located at the recovery stage of recall. Additional experimental work is needed to address the issue in more detail. In particular, this work should resolve the difference in effective retention interval between practiced items and control items to permit a more thorough comparison of the—first and subsequent—latencies of the two types of items. The results provide an answer on the question of the relation between facilitatory and inhibitory processes, which is a high priority for computational models of episodic recall.

Sampling/Recovery Versus Generate/Recognize Models

Results from prior work have demonstrated that recall total and response latency are empirically independent, with some experimental manipulations affecting recall total but not response latency (Metlay, Handley, & Kaplan, 1971; Rohrer, 1996; Rohrer & Wixted, 1994) and others affecting response latency but not recall total (Rohrer et al., 1999). This independence is of major theoretical significance as it is consistent with the view that recall is mediated by a two-stage process in which, in the first step, items are sampled from a set of to-be-recalled items and, in the second step, sampled items are recovered into consciousness if their absolute strength exceeds some threshold (Rohrer, 1996; see also Mensink & Raaijmakers, 1988; Raaijmakers & Shiffrin, 1981).

The current results support the independence of recall total and response latency by adding two further examples to the list of dissociations. As one dissociation, the results of Experiment 1 and Experiment 2 show a difference in recall total between unpracticed and control items but show no difference in response latency between the two types of items. As the other dissociation, the results of the control study of Experiment 1 show a strong differ-

ence in response latency between two testing conditions (presentation of category cues vs. subcategory cues) but reveal no effect on recall totals. The current results thus replicate both types of dissociations between recall total and response latency and support two-stage models of recall, in which a sampling and recovery stage are incorporated.

Another, very prominent class of two-stage models of recall is generate/recognize models. According to the classic version of these models, recall involves a generation stage in the first stage, in which candidates for recall are generated, and a subsequent recognition stage, in which each generated candidate is examined whether it was encountered in the original learning episode or not; if it was, then it is recalled (Kintsch, 1970; for a variant of the model, see also Jacoby & Hollingshead, 1990). Not much is yet known about the exact relationship between sampling/recovery and generate/recognize models.

A possible way to reconcile the two models is to assume that sampling and recovery are subprocesses of the generation stage of generate/recognize models. This assumption rests on the view that sampling and recovery both have to do with bringing information into consciousness, which is also the goal of the generation stage. Harmonizing the two types of two-stage models in this way suggests that inhibition in retrieval-induced forgetting acts at the generation stage of recall and not at the recognition stage. The contribution of the sampling/recovery approach to the problem of retrieval-induced forgetting thus can be regarded as a clue to get a more detailed picture on when exactly in the generation stage inhibition takes place. Further work on this point is needed.

Conclusions

In the current study, we addressed retrieval-induced forgetting by studying the time course of cued recall in this type of situation. Knowing how retrieval practice influences the dynamics of recall can improve the understanding of the nature of retrieval inhibition. The results show that retrieval practice affects recall total but has no influence on the recall dynamics. This finding suggests that retrieval practice reduces the activation level of unpracticed items and thus prevents them from exceeding the recovery threshold. Retrieval inhibition, therefore, should not operate at the sampling stage but rather at the recovery stage of recall. The results concerning the facilitatory effects of retrieval practice are similar and provide a first indication for the view that similar mechanisms are involved in the facilitatory and inhibitory effects of retrieval practice.

References

Anderson, M. C. (2003). Rethinking interference theory: Executive control and the mechanisms of forgetting. *Journal of Memory and Language, 49,* 415–445.

Anderson, M. C., & Bell, T. (2001). Forgetting our facts: The role of inhibitory processes in the loss of propositional knowledge. *Journal of Experimental Psychology: General, 130,* 544–570.

Anderson, M. C., Bjork, E. L., & Bjork, R. A. (2000). Retrieval-induced forgetting: Evidence for a recall-specific mechanism. *Psychonomic Bulletin & Review, 7,* 522–530.

Anderson, M. C., Bjork, R. A., & Bjork, E. L. (1994). Remembering can cause forgetting: Retrieval dynamics in long-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 1063–1087.

Anderson, M. C., & Green, C. (2001, March 15). Suppressing unwanted memories by executive control. *Nature, 410,* 366–369.

Anderson, M. C., & Spellman, B. A. (1995). On the status of inhibitory mechanisms in cognition: Memory retrieval as a model case. *Psychological Review, 102,* 68–100.

Basden, D. R., & Basden, B. H. (1995). Some tests of the strategy disruption hypothesis of part-list cuing inhibition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 1656–1669.

Battig, W. F., & Montague, W. E. (1969). Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology Monographs, 80*(3, Pt. 2), 1–46.

Bäuml, K.-H. (1998). Strong items get suppressed, weak items do not: The role of item strength in output interference. *Psychonomic Bulletin & Review, 5,* 459–463.

Bäuml, K.-H. (2002). Semantic generation can cause episodic forgetting. *Psychological Science, 13,* 357–361.

Bäuml, K.-H., & Aslan, A. (2004). Part-list cuing as instructed retrieval inhibition. *Memory & Cognition, 32,* 610–617.

Bäuml, K.-H., & Hartinger, A. (2002). On the role of item similarity in retrieval-induced forgetting. *Memory, 10,* 215–224.

Bäuml, K.-H., & Kuhbandner, C. (2003). Retrieval-induced forgetting and part-list cuing in associatively structured lists. *Memory & Cognition, 31,* 1188–1197.

Bousfield, W. A., & Sedgewick, C. H. W. (1944). An analysis of sequences of restricted associative responses. *Journal of General Psychology, 30,* 149–165.

Brown, R., & McNeill, D. (1966). The tip of the tongue phenomenon. *Journal of Verbal Learning and Verbal Behavior, 5,* 325–337.

Butler, K. M., Williams, C. C., Zacks, R. T., & Maki, R. H. (2001). A limit on retrieval-induced forgetting. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* 1314–1319.

Ciranni, M. A., & Shimamura, A. P. (1999). Retrieval-induced forgetting in episodic memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 1403–1414.

Conway, A. R. A., Tuholski, S. W., Shisler, R. J., & Engle, R. W. (1999). The effect of memory load on negative priming: An individual differences investigation. *Memory & Cognition, 27,* 1042–1050.

Dunn, E. W., & Spellman, B. A. (2003). Forgetting by remembering: Stereotype inhibition through rehearsal of alternative aspects of identity. *Journal of Experimental Social Psychology, 39,* 420–433.

Geiselman, R. E., Bjork, R. A., & Fishman, D. L. (1983). Disrupted retrieval in directed forgetting: A link with posthypnotic amnesia. *Journal of Experimental Psychology: General, 112,* 58–72.

Gronlund, S. D., & Shiffrin, R. M. (1986). Retrieval strategies in recall of natural categories and categorized lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12,* 550–561.

Herrmann, D. J., & Pearle, P. M. (1981). The proper role of clusters in mathematical models of continuous recall. *Journal of Mathematical Psychology, 24,* 139–162.

Hicks, J. L., & Starns, J. J. (2004). Retrieval-induced forgetting occurs in tests of item recognition. *Psychonomic Bulletin & Review, 11,* 125–130.

Hogan, R. M., & Kintsch, W. (1971). Differential effects of study and test trials on long-term recognition and recall. *Journal of Verbal Learning and Verbal Behavior, 10,* 562–567.

Jacoby, L. L., & Hollingshead, A. (1990). Toward a generate/recognize model of performance on direct and indirect tests of memory. *Journal of Memory and Language, 29,* 433–454.

Johansson, M., Aslan, A., Bäuml, K.-H., & Mecklinger, A. (2005). *When remembering causes forgetting: Electrophysiological correlates of retrieval-induced forgetting.* Unpublished manuscript.

Johnson, S. K., & Anderson, M. C. (2004). The role of inhibitory control in forgetting semantic knowledge. *Psychological Science, 15,* 448–453.

Kintsch, W. (1970). Models for free recall and recognition. In D. A.

Norman (Ed.), *Models of human memory* (pp. 331–373). New York: Academic Press.

Levy, B. J., & Anderson, M. C. (2002). Inhibitory processes and the control of memory retrieval. *Trends in Cognitive Sciences, 6,* 299–305.

MacLeod, C. M. (1998). Directed forgetting. In J. M. Golding & C. M. MacLeod (Eds.), *Intentional forgetting: Interdisciplinary approaches* (pp. 1–57). Mahwah, NJ: Erlbaum.

Mannhaupt, H.-R. (1983). Produktionsnormen für verbale Reaktionen zu 40 geläufigen Kategorien [Production norms for verbal reactions to 40 familiar categories]. *Sprache & Kognition, 2,* 264–278.

McGill, W. J. (1963). Stochastic latency mechanism. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (Vol. 1, pp. 309–360). New York: Wiley.

Mensink, J.-G., & Raaijmakers, J. G. W. (1988). A model of interference and forgetting. *Psychological Review, 95,* 434–455.

Metlay, W., Handley, A., & Kaplan, I. T. (1971). Memory search through categories of varying size. *Journal of Experimental Psychology, 91,* 215–219.

Morrison, D. G. (1979). An individual differences pure extinction process. *Journal of Mathematical Psychology, 19,* 307–315.

Moulin, C. J. A., Perfect, T. J., Conway, M. A., North, A. S., Jones, R. W., & James, N. (2002). Retrieval-induced forgetting in Alzheimer's disease. *Neuropsychologia, 40,* 862–867.

Nickerson, R. S. (1984). Retrieval inhibition from part-set cuing: A persisting enigma in memory research. *Memory & Cognition, 12,* 531–552.

Perfect, T. J., Moulin, C. J. A., Conway, M. A., & Perry, E. (2002). Accessing the inhibitory account of retrieval-induced forgetting with implicit-memory tests. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 1111–1119.

Perfect, T. J., Stark, L.-J., Tree, J. J., Moulin, C. J. A., Ahmed, L., & Hutter, R. (2004). Transfer appropriate forgetting: The cue-dependent nature of retrieval-induced forgetting. *Journal of Memory and Language, 51,* 399–417.

Postman, L., Stark, K., & Fraser, J. (1968). Temporal changes in interference. *Journal of Verbal Learning and Verbal Behavior, 7,* 672–694.

Raaijmakers, J. G. W., & Shiffrin, R. M. (1980). SAM: A theory of probabilistic search of associative memory. In G. Bower (Ed.), *The psychology of learning and motivation* (Vol. 14, pp. 207–262). New York: Academic Press.

Raaijmakers, J. G. W., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review, 88,* 93–134.

Roediger, H. L., III. (1978). Recall as a self-limiting process. *Memory & Cognition, 6,* 54–63.

Roediger, H. L., III, & Neely, J. H. (1982). Retrieval blocks in episodic and semantic memory. *Canadian Journal of Psychology, 36,* 213–242.

Roediger, H. L., III, & Schmidt, S. R. (1980). Output interference in the recall of categorized and paired associate lists. *Journal of Experimental Psychology: Human Learning and Memory, 6,* 91–105.

Roediger, H. L., III, Stellon, C. C., & Tulving, E. (1977). Inhibition from part-list cues and rate of recall. *Journal of Experimental Psychology: Human Learning and Memory, 3,* 174–188.

Rohrer, D. (1996). On the relative and absolute strength of a memory trace. *Memory & Cognition, 24,* 188–201.

Rohrer, D., Salmon, D. P., Wixted, J. T., & Paulsen, J. S. (1999). The disparate effects of Alzheimer's disease and Huntington's disease on semantic memory. *Neuropsychology, 13,* 381–388.

Rohrer, D., & Wixted, J. T. (1994). An analysis of latency and interresponse time in free recall. *Memory & Cognition, 22,* 511–524.

Rohrer, D., Wixted, J. T., Salmon, D. P., & Butters, N. (1995). Retrieval from semantic memory and its implications for Alzheimer's disease. *Journal of Experimental Psychology: Human Learning and Memory, 21,* 1127–1139.

Rundus, D. (1973). Negative effects of using list items as recall cues. *Journal of Verbal Learning and Verbal Behavior, 12,* 43–50.

Sahakyan, L., & Kelley, C. M. (2002). A contextual change account of the directed forgetting effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 1064–1072.

Scheithe, K., & Bäuml, K.-H. (1995). Deutschsprachige Normen für Vertreter von 48 Kategorien [German norms for members of 48 categories]. *Sprache & Kognition, 14,* 39–43.

Schooler, J. W., Fiore, S. M., & Brandimonte, M. A. (1997). At a loss from words: Verbal overshadowing of perceptual memoris. In D. L. Medin (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 37, pp. 291–340). San Diego, CA: Academic Press.

Shaw, J. S., III, Bjork, R. A., & Handal, A. (1995). Retrieval-induced forgetting in an eyewitness-memory paradigm. *Psychonomic Bulletin & Review, 2,* 249–253.

Smith, A. D. (1971). Output interference and organized recall from long-term memory. *Journal of Verbal Learning and Verbal Behavior, 10,* 400–408.

Starns, J. J., & Hicks, J. L. (2004). Episodic generation can cause semantic forgetting: Retrieval-induced forgetting of false memories. *Memory & Cognition, 32,* 602–609.

Tipper, S. P., & Cranston, M. (1985). Selective attention and priming: Inhibitory and facilitatory effects of ignored primes. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 37*(A), 591–611.

Tulving, E., & Arbuckle, T. Y. (1966). Input and output interference in short-term associative memory. *Journal of Experimental Psychology, 72,* 89–104.

Tulving, E., & Pearlstone, Z. (1966). Availability versus accessibility of information in memory for words. *Journal of Verbal Learning and Verbal Behavior, 5,* 381–391.

Tulving, E., & Psotka, J. (1971). Retroactive inhibition in free recall: Inaccessibility of information available in the memory store. *Journal of Experimental Psychology, 87,* 1–8.

Veling, H., & van Knippenberg, A. (2004). Remembering can cause inhibition: Retrieval-induced inhibition as cue independent process. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 315–318.

Vorberg, D., & Ulrich, R. (1987). Random search with unequal search rates: Serial and parallel generalizations of McGill's model. *Journal of Mathematical Psychology, 31,* 1–23.

Williams, C. C., & Zacks, R. T. (2001). Is retrieval-induced forgetting an inhibitory process? *American Journal of Psychology, 114,* 329–354.

Wixted, J. T., Ghadisha, H., & Vera, R. (1997). Recall latency following pure- and mixed-strength lists: A direct test of the relative strength model of free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 523–538.

Wixted, J. T., & Rohrer, D. (1993). Proactive interference and the dynamics of free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 1024–1039.

Wixted, J. T., & Rohrer, D. (1994). Analyzing the dynamics of free recall: An integrative review of the empirical literature. *Psychonomic Bulletin & Review, 1,* 89–106.

Zellner, M., & Bäuml, K.-H. (2004). Retrieval inhibition in episodic recall. In A. Mecklinger, H. Zimmer, & U. Lindenberger (Eds.), *Bound in memory: Insights from behavioral and neuropsychological studies* (pp. 1–26). Aachen, Germany: Shaker Verlag.

Zellner, M., & Bäuml, K.-H. (2005). Intact retrieval inhibition in children's episodic recall. *Memory & Cognition, 33,* 396–404.