Journal of Experimental Psychology:
Learning, Memory, and Cognition
2007, Vol. 33, No. 5, 863–875

0278-7393/07/$12.00 DOI: 10.1037/0278-7393.33.5.863

Retrieval-Induced Forgetting in Item Recognition: Evidence for a Reduction in General Memory Strength

Bernhard Spitzer and Karl-Heinz Bäuml
University of Regensburg

Retrieving a subset of previously studied material can impair later recognition of related items. Using the remember–know procedure (Experiment 1) and the receiver operating characteristic procedure (Experiment 2), the authors examined how such retrieval-induced forgetting can be explained in terms of single-process and dual-process accounts of recognition memory. Consistent across the 2 experiments, dual-process analysis suggested that retrieval practice reduces unpracticed items' familiarity but leaves their recollection largely unaffected, a finding that disagrees with prior work that points to recollective deficits in the forgotten items. Assuming that recognition is entirely based on a single source of memorial information, single-process analysis led to an excellent description of the data and suggested that retrieval practice reduces unpracticed items' general memory strength. This suggestion is consistent with prior work on free recall, cued recall, associative recognition, and response latencies and agrees with the inhibitory account of retrieval-induced forgetting. The authors argue that retrieval-induced forgetting in item recognition is caused by a reduction in general memory strength.

Keywords: episodic memory, retrieval-induced forgetting, recognition, recollection, familiarity

Retrieving a subset of formerly studied material can cause subsequent forgetting of the nonretrieved material. Such retrieval-induced forgetting has repeatedly been demonstrated with the retrieval-practice paradigm (Anderson, Bjork, & Bjork, 1994; for a review, see Anderson, 2003). In this paradigm, participants study items from different semantic categories (e.g., *FRUIT-orange*, *FRUIT-apple*, *INSECT-bee*).[1] Then, in a subsequent practice phase, they repeatedly retrieve half of the items from half of the studied categories using a word stem completion task (e.g., *FRUIT-or—*). After a distractor task, recall performance for all initially studied items is tested with a cued recall procedure in which the category name of the items is provided as a retrieval cue. The typical result in this experiment is that, relative to the control items from the unpracticed categories (e.g., *bee*), recall of the practiced material (*orange*) is improved and recall of the unpracticed material (*apple*) is impaired.

Retrieval-induced forgetting is a retrieval-specific effect (Anderson, Bjork, & Bjork, 2000; Bäuml, 2002; Ciranni & Shimamura, 1999). It has been observed with verbal (e.g., Anderson et al., 1994) and visual material (Ciranni & Shimamura, 1999) and has been demonstrated in a variety of settings, such as fact learning (Anderson & Bell, 2001), eyewitness memory (e.g., Saunders & MacLeod, 2002; Shaw, Bjork, & Handal, 1995), false memories (e.g., Bäuml & Kuhbandner, 2003; Starns & Hicks, 2004), and social cognition (e.g., Dunn & Spellman, 2003). It has mostly been observed in young adults but is present in young children (Zellner & Bäuml, 2005) and older adults (Aslan, Bäuml, & Pastötter, 2007) as well.

It is widely assumed that retrieval-induced forgetting is caused by retrieval inhibition. The proposal is that during retrieval practice on a subset of studied material, related unpracticed items interfere. To reduce this interference, the unpracticed material is inhibited, leading to persistent deactivation of the unpracticed items' memory representation (Anderson & Spellman, 1995; for a review, see Anderson, 2003; for noninhibitory accounts, see Perfect et al., 2004, or Williams & Zacks, 2001). Consistently, retrieval-induced forgetting has been found across a wide range of memory tests, including free recall, category-cued recall, and initial-letter cued recall (e.g., Anderson et al., 1994; Anderson & Spellman, 1995; Macrae & MacLeod, 1999). It has also been demonstrated in tests that employ so-called independent probes, that is, probes that were not used in a previous phase of the experiment (e.g., Anderson & Bell, 2001; Veling & Van Knippenberg, 2004) and in tests assessing recognition memory (Gómez-Ariza, Lechuga, Pelegrina, & Bajo, 2005; Hicks & Starns, 2004; Verde, 2004).

Dual-Process Accounts of Recognition Memory

Theories of recognition memory often assume that two distinct memory processes can independently contribute to successful recognition of studied material (for a review, see Yonelinas, 2002).

Bernhard Spitzer and Karl-Heinz Bäuml, Department of Experimental Psychology, University of Regensburg, Regensburg, Germany.

This research was supported by German Research Foundation Grant FOR 448 to Karl-Heinz Bäuml. We thank K. Nuissl and R. Schmidtner for their help in the experimental work and J. Dunn and N. Macmillan for their comments on a previous version of the article.

Correspondence concerning this article should be addressed to Karl-Heinz Bäuml, Department of Experimental Psychology, University of Regensburg, 93040, Regensburg, Germany. E-mail: karl-heinz.baeuml@psychologie.uni-regensburg.de

[1] Although usually the presentation styles for read versus spoken category cues and word items are different, in this article they will be given in a uniform style for consistency.

On the one hand, individuals can base their recognition judgments on conscious recollection of information about the spatiotemporal context of a studied episode. On the other, they can assess their subjective degree of stimulus familiarity in the absence of any recollective details, which is consistent with daily life experience, in which we may have the feeling of knowing a person without being able to remember where or when we originally met her. Although recollection is often conceptualized as a relatively slow process yielding qualitative information about previous events, familiarity reflects a relatively fast, strengthlike memory signal yielding purely quantitative information (Yonelinas, 2002).

In the literature, a number of behavioral methods have been suggested to measure the differential contribution of recollection and familiarity in recognition memory. Among the most prominent ones are the remember–know procedure (Tulving, 1985; Yonelinas & Jacoby, 1995) and the receiver operating characteristic (ROC) procedure (Yonelinas, 1994).

In the remember–know method, participants are asked to introspectively refine the qualitative nature of their subjective recognition experience when judging an item as old. They are instructed to respond "remember" whenever they dispose of any recollective details about the study event and to respond "know" in the absence of such information. To estimate recollection and familiarity processes from remember–know data, Yonelinas and Jacoby (1995) proposed the independence remember–know method, assuming that the proportion of remember responses indexes recollection ("remember" = R), whereas the proportion of know responses is thought to reflect familiarity in the absence of recollection, which can be expressed as "know" = $F(1 - R)$. The independence remember–know method allows for estimation of independent parameters for recollection and familiarity from the mutually exclusive remember–know response categories. Though the process-pure theoretical interpretation of the remember response as an index of recollection has been subject to criticism recently (e.g., Rotello, Macmillan, Reeder, & Wong 2005), the remember–know method is still widely accepted as an easily applicable way to yield at least a rough separation of recollective processes from overall recognition performance.

The proportions of correct and incorrect "old" responses to studied items (hit rate) and to new items (false alarms) are examined under different response criteria with the ROC procedure. Varying the response criterion is usually achieved by asking participants to rate their confidence of an item being old or new and cumulating the rating data across confidence levels. The cumulative false alarm and hit rates can then be plotted in x/y-space to obtain ROCs describing recognition performance. Empirical ROC curves obtained in recognition experiments usually are asymmetrical in shape, with higher discrimination performance at high confidence levels. Yonelinas (1994) interpreted this asymmetry as the result of the joint contribution of recollection and familiarity: Whereas familiarity alone is conceptualized as a strengthlike signal-detection process that yields a symmetrical ROC, recollection is thought to be a thresholdlike process, contributing a proportion of hits that is invariant across confidence levels. In the dual-process framework, the symmetrical familiarity-based ROC and the probability for recollection add up to the asymmetrical recognition ROC that is characterized by an upward shift of the curve at high confidence regions. Both processes can be formalized mathematically, and numerical estimates for recollection (R) and familiarity-based discrimination (d') can be derived from empirical ROC data (for details, see the Appendix).

Retrieval-Induced Forgetting and Dual-Process Accounts

Because recollection is often assumed to resemble recall-like processes, theoretically any variable affecting recall performance should also have an effect on the recollection component of recognition memory. Retrieval practice has been shown to impair cued recall and free recall of unpracticed items (for a review, see Anderson, 2003). Retrieval practice, therefore, should affect recollection-based recognition of unpracticed material as well.

Direct support for this proposal comes from a recent study by Verde (2004), who examined the detrimental effects of retrieval practice on unpracticed material using an associative-recognition test. In his experiment, pairs of semantically unrelated words were presented for associative study. A subset of these item–item pairs was subsequently retrieval practiced by providing one of the items as retrieval cue for recall of the associate. Finally, participants' associative memory was tested by presenting intact and rearranged pairs in a remember–know recognition test. The results showed significantly reduced discrimination performance for the unpracticed material at the remember criterion, suggesting a decrease in recollective processes. In contrast, no reliable forgetting at the old–new criterion (comprising both remember and know responses) showed up, indicating that the familiarity process was not reduced in this experiment.

Although, on the basis of dual-process models, Verde's (2004) work provides clear evidence that recollection can be disrupted in retrieval-induced forgetting, it is silent about whether familiarity plays a role in retrieval-induced forgetting as well. Indeed, there is broad evidence that associative recognition is largely based on recollection (Hockley & Consoli, 1999; Humphreys, 1978; Rotello, Macmillan, & Van Tassel, 2000) and thus may not be very susceptible to manipulations thought to target familiarity (Verde & Rotello, 2004). As a result, recognition memory in this previous study may have been mainly recollection-driven, so that a possible detrimental effect of retrieval practice on the familiarity of the unpracticed material was overlooked.

On the basis of previous studies, such a detrimental effect of retrieval practice on the familiarity process appears well possible. Veling and Van Knippenberg (2004), for instance, demonstrated retrieval-induced forgetting in a speeded-recognition task in which participants were asked to respond to recognition probes as fast as they could. Because it is generally assumed that recognition performance relies more on familiarity than on recollective processes when participants are required to make recognition decisions very quickly (e.g., Yonelinas, 2002), this finding may reflect more an effect of familiarity than of recollection.

Moreover, if retrieval-induced forgetting is the result of retrieval inhibition and leads to a weakening of the unpracticed items' memory representation (Anderson, 2003; Anderson & Spellman, 1995), then not only recollection of unpracticed items but also their familiarity may be reduced, given that familiarity is thought to reflect an item's general memory strength. In fact, recent work reported reliable retrieval-induced forgetting in old–new recognition performance (Gómez-Ariza et al., 2005; Hicks & Starns, 2004), which is at least consistent with a possible role of familiarity in this type of forgetting.

Retrieval-Induced Forgetting and Single-Process Accounts of Recognition Memory

In recent years, the dual-process framework has been subject to considerable criticism. In particular, it has been shown that the results from most recognition studies are compatible with a single-process account of recognition memory (e.g. Dunn, 2004; Wixted, 2007; Wixted & Stretch, 2004). According to the single-process view, recognition memory is entirely based on a single source of memorial information, that is, the memory strength of studied items relative to unstudied items. Recognition performance thus can be characterized by a standard signal-detection process assuming a single underlying dimension of memory strength, not unlike the dual-process model's conceptualization of stimulus familiarity. In contrast to dual-process models, single-process models exclude the independent contribution of a thresholdlike recollection process to recognition performance. Rather, to account for the characteristic shape of recognition ROCs, researchers often assume that the variance of the strength distribution for studied items can exceed the variance of the distribution for unstudied items (unequal-variance signal-detection model; for details, see the Appendix).

Single-process signal-detection models often lead to a more adequate description of empirical remember–know and ROC data than do dual-process accounts (e.g., Dunn, 2004, or Wixted, 2007). The same may hold when applying single-process models to recognition data from retrieval-induced forgetting experiments. If so, the effect of retrieval practice could be captured by modulations in the studied items' general memory strength. Retrieval practice should enhance the memory strength of practiced material, reflected in an increase of sensitivity (d'), and reduce that of unpracticed material, reflected in a decrease of sensitivity. Such a pattern would be consistent with the inhibitory account of retrieval-induced forgetting, which claims that retrieval practice weakens the memory representation of unpracticed items (Anderson, 2003; Anderson & Spellman, 1995). In particular, it would offer a more parsimonious account of the detrimental effects of retrieval practice on recognition memory than is suggested on the basis of the dual-process view.

Goals of the Study

Prior work showed that retrieval-induced forgetting occurs in recognition tests (Hicks & Starns, 2004) and indicated that retrieval practice affects unpracticed items' recollective processes (Anderson, 2003; Verde, 2004). The first goal of the present study was to investigate, on the basis of dual-process accounts of recognition memory, whether retrieval practice can affect the familiarity of the unpracticed material as well. Two experiments are reported, in each of which participants studied categorized material. After subjects practiced retrieval on a subset of the material, a standard item-recognition task was employed. To investigate possible qualitative effects of retrieval practice on recognition memory, we applied the remember–know procedure in Experiment 1 and the ROC procedure in Experiment 2. We expected converging results across the two measurement methods, thus providing an answer on whether retrieval practice impairs only recollection of unpracticed material or affects its familiarity as well. The second goal of the present study was to compare dual-process and single-process accounts of recognition memory and to determine which of the two accounts provides a better description of retrieval-induced forgetting in the present experiments. The results from this comparison will bear implications for theoretical interpretations of retrieval-induced forgetting, suggesting that retrieval practice affects unpracticed items' recollection and/or familiarity (dual-process view) or suggesting that it reduces unpracticed items' general memory strength (single-process view).

Experiment 1

Experiment 1 examines the detrimental effect of retrieval practice on item recognition by applying the remember–know procedure. According to dual-process accounts of the remember–know procedure, recognized items can be divided into a proportion of items that was recollected (*remember*) and a proportion of items that was not recollected (*know*). A possible decrease in recollection should lead to fewer remember responses, creating more know and/or more new responses; a decrease in familiarity should lead to fewer know and more new responses, and thus should lower overall recognition performance. Because retrieval practice has been shown to affect unpracticed items' recollection, we expected a decrease in remember responses for the unpracticed material. The crucial question was whether the expected decline in overall recognition performance would be the result of the expected reduction in remember performance or would be caused by a decrease in the unpracticed items' familiarity. Alternatively, the experiment may also yield results consistent with the single-process account of the remember–know procedure. In this case, any detrimental effect of retrieval practice on item recognition would be attributed to a decrease in the unpracticed items' general memory strength, thus contrasting with the dual-process view of a decrease in recollection and/or familiarity processes of unpracticed items.

Method

Participants. Subjects included 52 adults (18–40 years old), who took part in the experiment on a voluntary basis. They were tested individually in sessions that lasted approximately 30 min.

Materials. The study material was selected from associative lists that are typically employed in the Deese-Roediger-McDermott (DRM) paradigm when examining false memories of so-called critical items (Deese, 1959; Roediger & McDermott, 1995). Prior work reported retrieval-induced forgetting in Deese-Roediger-McDermott "categories" by showing that retrieval practice on half of the (studied) noncritical items can cause forgetting of the remaining (studied) noncritical items (Bäuml & Kuhbandner, 2003). In the present experiment, stimuli consisted of 12 exemplars from each of 8 target categories selected from the 24 Deese-Roediger-McDermott categories used by Roediger and McDermott (1995). The selected categories had as their critical items the words *DOCTOR, WINDOW, RIVER, COLD, SLOW, SLEEP, SWEET,* and *SOFT*, which served as the category cues in the experiment. All stimuli were idiomatically translated into German. The chosen exemplars were generally the 12 strongest associates to the critical item on the respective list. Additionally, two exemplars from each of the three categories *FRUIT, SMELL,* and *MOUNTAIN* were used as buffer items in the study phase, for a total of six buffer items.

Design. The experiment consisted of three main phases: a study phase, a retrieval-practice phase, and a final test phase. In the study phase, for each of the eight target categories the six intermediate associates of the critical item (ranks 4–9) were studied together with their critical item serving as category cue. The three strongest (ranks 1–3) and the three weakest associates (ranks 10–12) were used as lures in the final recognition test. In the retrieval-practice phase, from four of the eight studied categories, participants practiced the three weaker associates (ranks 7–9) whereas the three stronger associates (ranks 4–6) remained unpracticed. In this way, four types of target items were created: practiced weak associates (P+ items), unpracticed strong associates belonging to the same categories as the practiced items (P– items), and weak and strong associates from unpracticed categories serving as control items for the P+ and P– items (C+ and C– items, respectively). The design also created two types of new items in the final recognition test: lures from practiced categories (P lures) and lures from unpracticed categories (C lures). Across participants, we counterbalanced which of the studied categories were practiced.

Procedure. Participants were tested individually in a quiet surrounding. At the beginning of the study phase, participants were instructed to study each to-be-presented word together with its category cue (e.g., *DOCTOR-medicine, COLD-frosty,* etc.). The experimenter read out the 48 (8 × 6) experimental category–exemplar pairs one by one at a rate of 5 s per item. The serial order of the items was block randomized, that is, a random sequence of six blocks consisting of one randomly selected exemplar from each of the eight categories was presented to the participants with the constraint that no item in the sequence shared the next exemplar's category. Additionally, three of the six buffer items were shown at the beginning of the study list, and the remaining three were shown at the end. After half of the participants had taken the test, the order of the items was reversed. After the study phase, participants were instructed to count backward from 300 in steps of three for 60 s as a recency control.

In the subsequent retrieval-practice phase, the 12 (4 × 3) retrieval-practice cues were successively presented for word stem completion. In each of the 12 cases, the experimenter read out the category name together with the word stem of the to-be-recalled item (e.g., *COLD-fr—*) and noted the participant's response. After successful completion, or after 10 s without a response, the next item was read out. The order of the category–word stem pairs was block-randomized with the constraint that no item in the sequence shared the next exemplar's category. After half of the participants had taken the test, the order was reversed. Following the retrieval-practice phase, subjects were given a questionnaire and asked to rate German politicians on dimensions such as sympathy, competence, and honesty using Likert rating scales. They were warranted the anonymity of their information given.

In the final test phase, a sequential remember–know recognition test was conducted. Participants were instructed about the meaning of remember and know responses, illustrated by an example from everyday life (recognizing a person's face) and were informed about the basic testing procedure. Then the experimenter started reading aloud the test items. The category cues were not provided. The order of the test items was again block-randomized, with the additional constraint that no type of item requiring the same correct response (target or lure) appeared more than four times in a row. Again, after half of the participants had finished, the order was reversed. After presentation of an item, participants were asked to judge the item as old or new. After an "old" response, participants additionally had to give a "remember" or "know" judgment to further specify the quality of their remembering. In the case of a "new" response, the experimenter immediately proceeded with the next item. All responses were noted by the experimenter on a prepared data sheet.

Results

Retrieval-practice phase. In the retrieval-practice phase, participants, on average, successfully completed 75.7% (SE = 0.02) of the category–word stem pairs.

Recognition data. The recognition data are depicted in Table 1. Recognition and remember hit and false alarm rates were examined separately for practiced (P+) and unpracticed (P–) items.

The recognition hit rate was .85 for practiced (P+) items and .65 for their controls (C+), and the remember hit rate was .51 for practiced items and .34 for their controls. A 2 × 2 analysis of variance (ANOVA) with the two factors of item type (practiced vs. control) and criterion (old–new vs. remember) showed a significant main effect of item type, $F(1, 51) = 65.66$, $MSE = 0.026$, $p < .001$, partial $\eta^2 = .56$, and a significant main effect of criterion, $F(1, 51) = 91.72$, $MSE = 0.061$, $p < .001$, partial $\eta^2 = .64$. These effects reflect more hits for practiced items (P+) than for their controls (C+) and more frequent recognition than remember hits. No interaction between retrieval status and response criterion arose, $F(1, 51) = 1.47$, $MSE = 0.011$, $p > .20$.

The recognition hit rate was .60 for unpracticed items (P–) and .68 for their controls (C–), and the remember hit rate was .34 for unpracticed items and .37 for their controls. A 2 × 2 ANOVA with the two factors of item type (unpracticed vs. control) and criterion (old–new vs. remember) yielded a significant main effect of item type, $F(1, 51) = 5.07$, $MSE = 0.031$, $p < .05$, partial $\eta^2 = .09$, and a significant main effect of criterion, $F(1, 51) = 144.04$, $MSE =$

Table 1
Hit Rates and False Alarm Rates in Experiment 1

Item type	Hits		False alarms	
	Rate	*SE*	Rate	*SE*
	Old–new recognition performance			
P+	0.85[a]	0.02	0.22	0.02
C+	0.65	0.03	0.20	0.02
P−	0.60[a]	0.03	0.22	0.02
C−	0.68	0.02	0.20	0.02
	Remember performance			
P+	0.51[a]	0.04	0.07	0.02
C+	0.34	0.03	0.06	0.01
P−	0.34	0.03	0.07	0.02
C−	0.37	0.03	0.06	0.01

Note. P+ = practiced items; P− = unpracticed items; C+ = control items for the practiced items; C− = control items for the unpracticed items.
[a] Significant deviations from control performance ($p < .01$).

0.031, $p < .001$, partial $\eta^2 = .74$, reflecting fewer hits to unpracticed items (P–) than to their controls (C–) and again more frequent recognition than remember hits. The interaction between the factors of item type and response criterion was marginally significant, $F(1, 51) = 3.20$, $MSE = 0.008$, $p < .10$, partial $\eta^2 = .06$. Planned comparisons revealed no reliable reduction for remember hits, $t(51) = 1.20$, $p > .20$, but significantly fewer recognition hits for unpracticed material, $t(51) = 2.85$, $p < .01$, $d = 0.43$.

False alarm rates were examined separately. As expected, a 2 × 2 ANOVA with the two factors of category type (practiced vs. control) and criterion (old–new vs. remember) showed a significant main effect of criterion, $F(1, 51) = 125.930$, $MSE = 0.009$, $p < .001$, partial $\eta^2 = .71$, reflecting the fact that recognition false alarms are far more frequent than remember false alarms. Besides, no significant main effect of category type, $F(1, 51) = 1.46$, $MSE = 0.009$, $p > .25$, and no interaction between retrieval status and response criterion, $F(1, 51) = 1.64$, $MSE = 0.002$, $p > .30$, were found.

Dual-process analysis. For dual-process analysis of the data, the independence remember–know model by Yonelinas and Jacoby (1995) was applied to the remember–know raw data. Recollection was estimated as the proportion of remember responses to old items minus the proportion of remember responses to new items. Familiarity (F) was estimated for old and new items by dividing the proportion of know responses by the complement of the proportion of remember responses: Familiarity = Know/(1 – Remember). Then the familiarity estimate for the new items was subtracted from that for the old items (for details, see Yonelinas, 2002, or Yonelinas & Jacoby, 1995). The process estimates are depicted in Table 2.[2]

The recollection estimates were .44 for practiced items (P+) and .28 for their controls (C+). The familiarity estimates were .45 for P+ items and .29 for C+ items. A 2 × 2 ANOVA with the two factors of item type (practiced vs. control) and process estimate (recollection vs. familiarity) showed a significant main effect of item type, $F(1, 51) = 37.52$, $MSE = 0.033$, $p < .001$, partial $\eta^2 = .42$, indicating a general memory improvement for the practiced material. The estimates for recollection and familiarity did not differ significantly, $F(1, 51) < 1$, and there was no interaction between the two factors, $F(1, 51) < 1$. Planned comparisons revealed a significant increase in both recollection, $t(51) = 4.47$, $p < .001$, $d = 0.61$, two-tailed, and familiarity, $t(51) = 4.10$, $p < .001$, $d = 0.49$, two-tailed, of the practiced items.

The recollection estimate for unpracticed items (P–) was .27 compared with .31 for their controls (C–). The familiarity estimates were .23 for P– items and .31 for C– items. A 2 × 2 ANOVA with the two factors of item type (unpracticed vs. control) and process estimate (recollection vs. familiarity) showed a significant main effect of item type, $F(1, 51) = 5.36$, $MSE = 0.034$, $p < .05$, partial $\eta^2 = .10$, indicating generally impaired memory processes for the unpracticed material. The estimates for recollection and familiarity did not differ significantly, $F(1, 51) < 1$, and there was no interaction between the two factors, $F(1, 51) < 1$. Planned comparisons revealed a significant decrease in familiarity of the unpracticed material, $t(51) = 2.02$, $p < .05$, $d = 0.34$, but no reliable reduction in recollective processes, $t(51) = 1.31$, $p > .15$.

We also fitted the sum–difference theory of remembering and knowing (STREAK; Rotello, Macmillan, & Reeder, 2004) to the remember–know data. STREAK assumes two different continuous sources of memorial information: global memory strength reflecting stimulus familiarity and specific memory strength reflecting recollective information. Old–new discrimination is proposed to be based on a weighted sum of global and specific memory strength, whereas remember–know decisions are thought to be based on a weighted difference of the two sources of information. Given the standard deviation of the new items' strength distribution (s), the model yields separate parameter estimates for the diagnosticity of both global and specific memory strength (d_x and d_y, respectively). Following Rotello et al. (2004), parameter s was set to 0.8 for each item type.

The best-fitting STREAK parameters for both global and specific memory strength are shown in the lower panel of Table 2. Both parameters were higher for practiced items than for their controls (d_x: 1.00 vs. 0.64, d_y: 1.46 vs. 0.97). For unpracticed items, global as well as specific memory strength was lower than for their controls (d_x: 0.50 vs. 0.68, d_y: 0.90 vs. 1.05). With four free parameters to account for the four data points in the remember–know procedure (remember and know hits and false alarms), the STREAK model is saturated, and thus its goodness of fit cannot be tested statistically. However, likelihood-ratio tests can be used to examine whether the model's parameters vary with item type. For P+ items, the increase in global as well as in

Table 2
Dual-Process Parameter Estimates for Experiment 1

Item type	F	R	d_x	d_y	C_O	C_r
		Independence remember–know model				
P+	.45[a]	.44[a]				
C+	.29	.28				
P−	.23[a]	.27				
C−	.31	.32				
		STREAK model				
P+			1.00[a]	1.46[a]	0.61	0.24
C+			0.64	0.97	0.67	0.06
P−			0.50[a]	0.90	0.61	0.14
C−			0.68	1.05	0.67	0.10

Note. P+ = practiced items; P− = unpracticed items; C+ = control items for the practiced items; C− = control items for the unpracticed items; F = familiarity; R = recollection; d_x = global memory strength; d_y = specific memory strength; C_O = location of the sum–difference theory of remembering and knowing (STREAK) old/new criterion; C_r = location of the STREAK remember criterion.
[a] Significant deviations from control performance ($p < .05$).

[2] The retrieval-practice paradigm yields relatively few observations per participant and item type. Thus, fitting the data on an individual-participant level would cause major distortions of the models' parameter estimates. Accordingly, most modeling analysis in the present study was performed on aggregate data. The only exception was the data from the independence remember–know procedure, where, to allow for statistical testing, the parameters were calculated on an individual-participant level. Applying the independence remember–know model to the aggregate data, however, led to an identical pattern of results as the model's application to the individual-participant data.

specific memory strength was statistically reliable, d_x: $\chi^2(1) = 14.53$, $p < .001$; d_y: $\chi^2(1) = 19.36$, $p < .001$. For P– items, only d_x was significantly reduced, $\chi^2(1) = 5.05$, $p < .05$, whereas the decrease in d_y was not statistically reliable, $\chi^2(1) = 1.62$, $p > .20$.

Single-process analysis. To account for the data from the remember–know procedure, the single-process signal-detection model assumes a single parameter d' for the general memory strength of old items relative to new items, and response criteria for both remember and know responses (parameters r and k) ordered on the same strength continuum (e.g., Dunn, 2004; see also the Appendix). Assuming equal variances of the underlying normal distributions for old and new items, the single-process model has three parameters to fit the four data points in the remember–know procedure. Thus, the model has one degree of freedom for testing its goodness of fit. The equal-variance model described the data well for three of the four item types, all $\chi^2(1)$s $<$ 1.50, $p > .25$, but had to be rejected for P+ items, $\chi^2(1) = 8.39$, $p < .005$.

We also fitted the more general unequal-variance signal-detection model to the data, allowing the variance of the old items' distribution (parameter sigma) to vary freely. Analogous to the STREAK model reported above, with four free parameters to account for the four data points in the remember–know procedure, the unequal-variance signal-detection model is saturated, and thus its goodness of fit cannot be tested statistically. However, when directly compared with the more restrictive equal-variance model, the unequal-variance model described the data set significantly better, $\chi^2(4) = 10.59$, $p < .05$, and stable maximum-likelihood estimates for the model's parameters could be derived for each of the four item types (see lower panel of Table 3).

Likelihood-ratio tests were used to examine whether the parameters of the unequal-variance model varied across item type. The parameter estimates showed significantly higher *d*s for P+ items compared with C+ items (1.50 vs. 1.18), $\chi^2(1) = 16.10$, $p < .001$, indicating an increase in memory strength of the practiced material. They also showed significantly lower *d*s for P– items compared with C– items (1.04 vs. 1.25), $\chi^2(1) = 9.52$, $p < .005$, indicating a reduction in the unpracticed material's memory strength. The placement of remember and know criteria (parameters r and k) did not vary with item type, $\chi^2(6) = 5.40$, $p > .40$, nor did parameter sigma, $\chi^2(3) = 4.85$, $p > .15$.[3,4]

Table 3
Single-Process Parameter Estimates for Experiment 1

Item type	d'	σ	k	r
	Equal-variance signal-detection model			
P+[a]	1.65[b]	1.00	0.72	1.57
C+	1.19	1.00	0.83	1.57
P−	1.04[b]	1.00	0.76	1.48
C−	1.27	1.00	0.83	1.58
	Unequal-variance signal-detection model			
P+	1.50[b]	0.71	0.76	1.49
C+	1.18	0.88	0.84	1.54
P−	1.04[b]	1.05	0.76	1.49
C−	1.25	0.87	0.84	1.54

Note. P+ = practiced items; P− = unpracticed items; C+ = control items for the practiced items; C− = control items for the unpracticed items; d' = familiarity; σ = variance of the target distribution (set to 1.00 in the equal-variance model); k = location of the know criterion; r = location of the remember criterion.
[a] Insufficient goodness of fit. [b] Significant deviations from control performance ($p < .01$).

Discussion

With the help of a standard-item recognition task, Experiment 1 produced results that showed that retrieval practice impaired recognition of unpracticed items at the old–new criterion but had no reliable effect at the remember criterion. Consistently, dual-process analysis of the remember–know data indicated that retrieval practice lowered the unpracticed material's familiarity (global memory strength) but had only a small and nonreliable detrimental effect on the material's recollection (specific memory strength). Recollection is often assumed to resemble recall-like processes (e.g., Yonelinas, 2002). Any variable reducing recall thus should also reduce the recollection component of recognition memory. Because retrieval-induced forgetting has been reported in a large number of free and cued recall experiments, following this rationale, one should also expect impaired recollective processes in the unpracticed material's recognition. The present finding of no reliable recollective impairment is inconsistent with this expectation and thus challenges a dual-process account of the present results.

In contrast to the dual-process accounts, the single-process signal-detection model assumes that recognition memory is entirely based on a single source of memorial information. Applying this model to the remember–know data of the present experiment indicates that the forgetting of the unpracticed items was caused by a decrease in the items' general memory strength. This view is consistent with the inhibitory account of retrieval-induced forgetting, according to which retrieval practice weakens the unpracticed material's memory representation and, due to the items' reduced

[3] When applying the unequal-variance signal-detection model to recognition data, it is usually assumed that the variance of the strength distribution for studied items can exceed the variance of the unstudied items' distribution (i.e., $\sigma \geq 1$). In contrast to this assumption, the estimates for sigma in Experiment 1 tended to be smaller than 1. The results from a recent meta-analysis (Rotello et al., 2004), however, showed that average remember–know data often yield two-point zROCs with slopes larger than 1 (and thus $\sigma < 1$; for a related demonstration, see Malmberg & Xu, 2006). The relatively small values for sigma in Experiment 1 thus are most likely due to the requisite analysis of aggregate data rather than being indicative of inconsistencies with the single-process model.

[4] Like the dual-process models, the single-process models were fit separately to each of the single item types. This analysis yields separate estimates of the false alarm rate for each item type. Because all types of items were presented randomly in the same test phase (see Method section), arguably participants may have adopted the same decision criteria for the single-item types, suggesting that false alarm rates may be constrained to be equal across item type (e.g., Stretch & Wixted, 1998). We therefore reanalyzed the data by fitting the signal-detection models simultaneously to all six item types (P+, P–, C+, C–, P_{new}, C_{new}). Despite some minor differences in absolute values of the parameter estimates, the pattern of results was largely indistinguishable from the pattern obtained when fitting the signal-detection models separately to the single-item types. The results thus do not depend on type of analysis.

memory strength, should impair free recall, cued recall, and recognition of the items (Anderson, 2003; Anderson & Spellman, 1995). The single-process view also provides a more parsimonious account of the data than does the dual-process view.

Although the results of Experiment 1 suggest a preference for the single-process interpretation of retrieval-induced forgetting, in general, such a preference might be premature. Indeed, because standard remember–know data are not rich enough to allow for statistical testing of single-process and dual-process models, the data of Experiment 1 did not allow a statistical evaluation of the two models. Therefore, in Experiment 2 we used ROC methods to examine retrieval-induced forgetting. The data from ROC methods are rich enough to allow for statistical testing and thus may offer an additional source of information for evaluating the two types of models.

Experiment 2

With the help of the dual-process framework, Experiment 1 yielded converging results across analysis methods regarding the detrimental effects of retrieval practice on recollection and familiarity, suggesting that retrieval practice reduces mainly unpracticed items' familiarity and hardly, if at all, their recollection. Addressing the generalizability of this result, Experiment 2 examines whether such converging evidence is also present across different measurement procedures. In Experiment 2, we therefore used the ROC procedure to further examine the effects of retrieval practice on recollection and familiarity processes. Also, using this procedure, both the dual-process and the single-process models' ability to account for the recognition data can be tested statistically. These data will help evaluate which of the two types of models provides a better account of retrieval-induced forgetting in item recognition.

Method

Participants. Participants were 48 young adults (19–31 years old), who were paid 5 Euro for taking part in the experiment. They were tested individually in sessions that lasted approximately 35 min.

Materials. Stimuli were 12 concrete German words from each of eight semantic categories (Battig & Montague, 1969; Mannhaupt, 1983). The categories were *BODY PART, SPORT, MUSICAL INSTRUMENT, QUADRUPED, PIECE OF FURNITURE, TOOL KIT, SPICE*, and *TREE*. The chosen exemplars were 12 strong and moderate exemplars of the respective category (ranks 4–15 in the norms). Within each of the eight categories, six of the chosen exemplars (ranks 7–12) were studied, and the remaining six (ranks 4–6 and 13–15) were used as lures in the recognition test. Additionally, five exemplars from each of the three categories *COLOR, RELATIVE,* and *ROAD SIGN* were used as buffer items in the study list, for a total of 15 buffer items.

Design. The experimental design was identical to that in Experiment 1, with the exception that retrieval practice was not restricted to weaker exemplars. For four of the eight target categories, participants practiced retrieval of three of the six studied exemplars, with each item of a category serving equally often as a practiced and unpracticed item across participants. Correspondingly, across participants each item from unpracticed categories served equally often as a control for practiced and unpracticed items. In contrast to Experiment 1, control items thus were not divided into weak and strong exemplars. As a result, the experiment yielded three types of target items: practiced items (P+ items), unpracticed items belonging to the same categories as the practiced items (P– items), and unpracticed items from unpracticed categories, which served as control items (C items). Across participants, we counterbalanced which categories were retrieval-practiced.

Procedure. Participants were tested individually in a quiet surrounding, seated in front of a 15-in. computer screen. At the beginning of the study phase, an instruction to study all to-be-presented words was displayed. Then, each of the 48 (8×6) experimental items was presented for 2,500 ms, followed by a 500-ms blank screen. In contrast to Experiment 1, the items were presented without their category cue, which was done to minimize the possible role of contextual cues at encoding (see Hicks & Starns, 2004). The order of the items was randomized in the same manner as in Experiment 1. Additionally, six buffer items were shown at the beginning of the study list, and six were shown at the end. Directly after the study phase, participants were given a prepared sheet containing arithmetic problems and were instructed to solve as many problems as possible within 2 min.

Subsequently, the retrieval-practice phase started in which participants were asked to practice 12 (4×3) previously studied items. A category–word stem pair was presented on the screen (e.g., *FRUIT-ap—*), and participants were instructed to complete the word stem with a studied item. The experimenter noted the participants' response on a prepared data sheet, and participants could proceed to the next item by pressing a key. The order of the category–word stem pairs was block-randomized in the same manner as in Experiment 1. For the following 7.5 min, a series of decision problems was presented, with each problem requiring a decision between two risky choices (Kahneman, Slovic, & Tversky, 1982).

In the final test phase, an old–new recognition test was applied in which participants were asked to rate their confidence of an item being old or new using a six-point rating scale ranging from *1: Definitely old* to *6: Definitely new*. Participants were encouraged to use the whole range of the rating scale throughout the testing procedure to specify their degree of confidence as accurately as possible. Each test item was presented in the middle of the screen, together with a schematically depicted rating scale in the lower part of the screen. Participants were instructed to enter their responses via the digits on the PC keyboard. As soon as any numerical response was entered, the next item was presented on the screen. The order of the items was block-randomized in the same manner as in Experiment 1. After a few practice trials, all test items were presented, and the participants' responses were recorded automatically in a log file.

Results

Retrieval-practice phase. On average, the participants successfully completed 87.8% ($SE = 0.02$) of the category–word stem pairs presented during retrieval practice.

Recognition data. For ROC analysis of the data, hit and false alarm rates were cumulated over the five criterion points of the confidence rating scale, starting with the most confident criterion point (*Definitely old*). The ROCs obtained by plotting the cumulative false alarm rates against the hit rates for each item type are illustrated in Figure 1.

Dual-process analysis. We fitted the dual-process signal-detection model (Yonelinas, 1994) to the raw data, assuming a thresholdlike recollection process in addition to a strengthlike signal–detection-familiarity process (for details, see the Appendix). When applied to 5-point ROC data, the model has seven free parameters (familiarity d', recollection R, and response criteria c_1–c_5) to fit the 10 data points in an old–new rating experiment (hits and false alarms for confidence levels 1–5). Thus, the model has three degrees of freedom for testing its goodness of fit. The model parameters were estimated with maximum-likelihood techniques, which also allow for statistical testing.

The dual-process model described the data for P+ and P– items well, P+: $\chi^2(3) = 2.92$, $p > .40$; P–: $\chi^2(3) = 1.38$, $p > .70$, but could not satisfactorily fit the data for the C items, $\chi^2(3) = 9.14$, $p < .05$. Still, stable maximum-likelihood parameter estimates could be derived for each item type. The parameter estimates for R and d' are depicted in Table 4. The estimates for d' were 1.65 for P+ items, 0.80 for P– items, and 1.10 for C items. The estimates for R were 0.67 for P+ items, 0.38 for P– items, and 0.40 for C items.

Although the model failed to fit the control items satisfactorily, likelihood-ratio tests were conducted to examine whether the three types of items differed in their recollection and familiarity estimates. Regarding the beneficial effects of retrieval practice, the analysis showed significantly higher performance for P+ items than for C items with respect to both recollection as measured by parameter R, $\chi^2(1) = 4.23$, $p < .05$, and familiarity as measured by parameter d', $\chi^2(1) = 5.65$, $p < .05$. Regarding the detrimental effects of retrieval practice, a significant difference between P– items and C items emerged with respect to their familiarity, d': $\chi^2(1) = 4.57$, $p < .05$, but not with respect to their recollective processes, R: $\chi^2(1) = 0.05$, $p > .95$.

Single-process analysis. As in Experiment 1, we fitted the equal-variance signal-detection model to the data. When applied to 5-point ROC data, this restrictive version of the single-process model has six free parameters (sensitivity d' and five response criteria) to fit the 10 data points in an old–new rating experiment. The equal-variance model could not describe the recognition ROCs and had to be rejected for all three item types, all $\chi^2(4)$s $>$ 14.00, all ps $< .01$. We therefore fitted the more general unequal-variance signal-detection model to the data. When applied to 5-point ROC data, the unequal-variance signal-detection model has seven free parameters (sensitivity d', target variance sigma, and five response criteria) and thus three degrees of freedom for statistically testing its goodness of fit. The unequal-variance signal-detection model described the data of all three item types well, all $\chi^2(3)$s < 4.00, all ps $> .30$. The goodness-of-fit statistics and the maximum-likelihood parameter estimates are depicted in Table 4.

Likelihood-ratio tests were used to examine whether the unequal-variance model's parameters varied with item type. The analysis showed that sensitivity as measured by d' was significantly higher for P+ items than for C items (3.03 vs. 1.85), $\chi^2(1) = 43.73$, $p < .001$, suggesting an increase in memory strength for practiced material. Also, parameter d' was reliably lower for P– items than for C items (1.58 vs. 1.85), $\chi^2(1) = 7.24$, $p < .01$, suggesting a significant reduction in memory strength for unpracticed items. Neither the variance of the old items' distribution as estimated by parameter sigma nor the placement of the five confidence criteria varied reliably across item type, $\chi^2(2) = 1.53$, $p > .40$; $\chi^2(10) = 5.83$, $p > .80$.

Discussion

With the help of the remember–know procedure, Experiment 1 produced results that showed reliable retrieval-induced forgetting and, on the basis of the dual-process view, suggested that retrieval practice mainly reduced the unpracticed items' familiarity. With

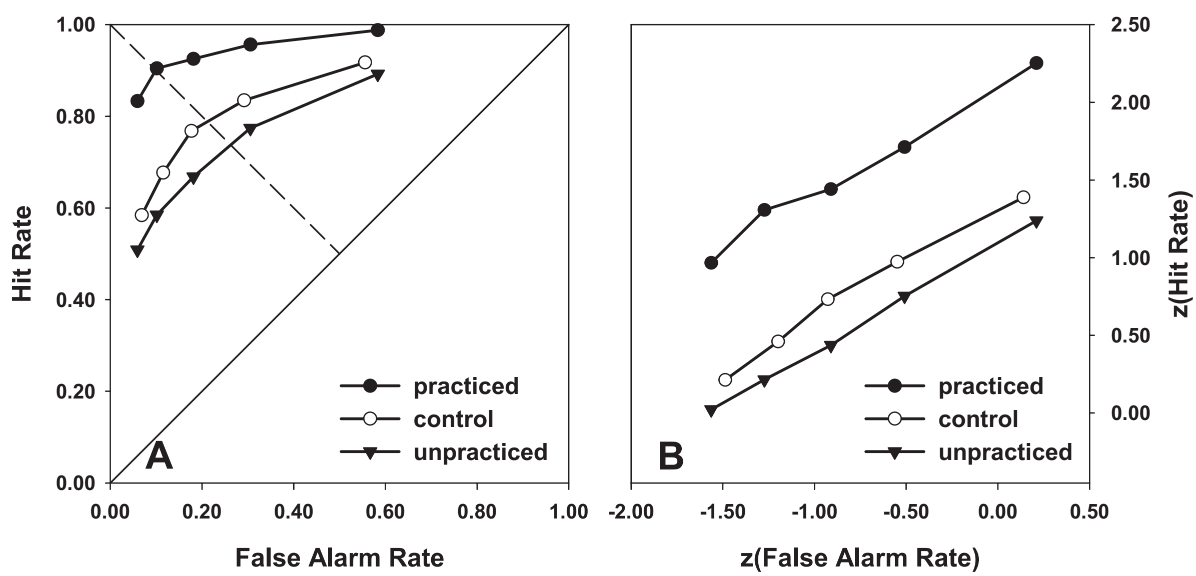

Figure 1. (A) Recognition receiver operating characteristics (ROCs) and (B) zROCs for the three item types in Experiment 2.

Table 4
Dual-Process and Single-Process Parameter Estimates for Experiment 2

	Parameter estimates			Goodness of fit		
Item type	d'	R	σ	χ^2	df	p
Dual-process signal-detection model						
P+	1.65[a]	0.67[a]		2.92	3	.40
C	1.10	0.40		9.14[b]	3	.03
P−	0.80[a]	0.38		1.38	3	.71
Unequal-variance signal-detection model						
P+	3.03[a]		1.45	1.95	3	.58
C	1.85		1.38	3.59	3	.31
P−	1.58[a]		1.45	0.26	3	.97

Note. P+ = practiced items; P− = unpracticed items; C = control items; d' = familiarity (dual-process model) or general memory strength (unequal-variance model); R = recollection; σ = variance of the target distribution.
[a] Significant deviations from control performance ($p < .05$). [b] Insufficient goodness of fit.

the help of the ROC procedure, Experiment 2 again found a detrimental effect of retrieval practice on overall recognition performance, well comparable in size to the forgetting observed in Experiment 1 at the old–new criterion. More important, dual-process analysis of the ROC data again indicated a decrease in the unpracticed items' familiarity but no reliable reduction in recollection. Dual-process analysis of remember–know data and dual-process analysis of ROC data thus converge on the view that, in item recognition, retrieval practice impairs mainly unpracticed items' familiarity.

The dual-process model described the data for the practiced and unpracticed items well but failed to describe the data for the control items. It thus did worse than the unequal-variance single-process model, which accounted for the data of all three item types and, in all three cases, led to better goodness-of-fit statistics than did the dual-process account. On the basis of the single-process view, the result of a decrease in sensitivity of the unpracticed items indicates that retrieval practice reduced the general memory strength of the unpracticed material. This interpretation of the ROC data of Experiment 2 agrees with the interpretation of the remember–know data of Experiment 1, converging on a common interpretation of the detrimental effects of retrieval practice. Because of the dual-process model's failure to find a reliable recollective deficit in unpracticed items and the single-process model's excellent description of the ROC data, Experiment 2, like Experiment 1, suggests a preference for the single-process interpretation of the data.

General Discussion

Dual-Process Account of the Data

On the basis of the dual-process account of recognition memory (e.g., Yonelinas, 2002), the results from prior free recall, cued recall, and associative-recognition studies indicated that retrieval practice on a subset of studied material can have detrimental effects on the unpracticed material's recollection (Anderson, 2003; Verde, 2004). One of the goals of the present study was to investigate whether, in light of the dual-process view, retrieval-induced forgetting is purely recollection-driven or whether there is an additional role of familiarity in this type of forgetting. Experiment 1 addressed the issue by using the remember–know procedure, Experiment 2 by using the ROC procedure.

The two experiments replicated the finding that retrieval-induced forgetting occurs in recognition tests (Gómez-Ariza et al., 2005; Hicks & Starns, 2004; Verde, 2004). More important, in both experiments the forgetting was accompanied by a reliable reduction in the unpracticed items' familiarity. In both experiments, however, the forgetting was not accompanied by a reliable reduction in recollection. This finding suggests that retrieval practice mainly impaired the familiarity of the unpracticed material and left its recollection largely unaffected. This pattern comes as a surprise and provides a challenge for the dual-process view of recognition memory.

The difference between the present and prior work in the suggested role of recollection in retrieval-induced forgetting might be due to differences in methodology across experiments. For instance, Verde (2004) reported reduced recollection and unaffected familiarity in retrieval-induced forgetting, which contrasts with the present finding. In the study by Verde, however, individuals were given pairs of semantically unrelated words for associative study, whereas in the present experiment single items had to be remembered. Such a difference might affect results, leading to different effects on unpracticed items' recollection (but see below for an alternative explanation). However, with the exception of the use of a recognition test, the present experiments used similar material and procedure as was used in the prior free and cued recall work (e.g., Bäuml & Kuhbandner, 2003, 2007). The difference in findings across these studies, therefore, cannot easily be attributed to methodological differences. The pattern of affected recollection in free and cued recall and unaffected recollection in item recognition thus provides a challenge for the dual-process view.

Single-Process Account of the Data

The data of the present experiments were well described by the unequal-variance variant of the single-process signal-detection model (e.g., Dunn, 2004; Wixted & Stretch, 2004). Notably, in Experiment 2 the model could describe the data of all three item types and thus did better than did the dual-process model, which failed to account for the data of one item type. These findings support a single-process view, according to which recognition memory in the present experiments relied on a single source of memorial information. Following this view, the results suggest that retrieval practice reduced the general memory strength of the unpracticed items. This holds both in Experiment 1 when applying the remember–know procedure and in Experiment 2 when applying the ROC procedure. The single-process account thus provides a consistent explanation of retrieval-induced forgetting in the present experiments.

The present explanation of retrieval-induced forgetting in terms of a reduction in the unpracticed items' general memory strength agrees with the results from prior free and cued recall experiments. The finding of retrieval-induced forgetting in free and cued recall experiments has been explained in terms of inhibition. According to this account, not-to-be-practiced items interfere during retrieval practice, and inhibition reduces their interference potential by weakening the items' memory representation. As a result of this inhibition, nonretrieved items behave like items with relatively weak memory representations and show reduced performance across a wide range of memory tests, including free recall, cued recall, and also item recognition (e.g., Anderson, 2003; Hicks & Starns, 2004).

The proposal that retrieval practice impairs unpracticed items' general memory strength is also consistent with the results from a recent response latency study. Response latency analysis casts light on the dynamics of recall, allowing conclusions about the size of the underlying search set and the memory strength of the set's items (for a review, see Wixted & Rohrer, 1994). Applying response latency analysis to retrieval-induced forgetting, Bäuml, Zellner, and Vilimek (2005) found that retrieval practice reduces unpracticed items' recall probability but does not affect their response latency. This result mirrors typical effects of item strength manipulations as they occur as a result of variations in study time or study trials (Rohrer, 1996; Wixted, Ghadisha, & Vera, 1997). In particular, it agrees with the suggested single-process view of retrieval-induced forgetting, according to which retrieval practice reduces the memory strength of unpracticed material.

Relation to Prior Associative-Recognition Work

On the basis of the dual-process view of recognition memory, the outcomes of the present study contrast with the results by Verde (2004). Whereas in the present item-recognition experiments unpracticed items showed a reliable deficit in familiarity but not in recollection, in Verde's associative-recognition experiment the opposing pattern arose with a reduction in unpracticed items' recollection but not in their familiarity. Verde drew his conclusions from a mainly descriptive analysis of the data without fitting dual-process or single-process models. We therefore reanalyzed the data of his Experiment 1, applying dual-process and single-process analysis as used in the present study.

When applying the independence remember–know model and the STREAK model to the aggregate data of Verde's (2004) experiment, the two dual-process analysis methods converged on showing a stronger decline in recollection (specific memory strength) than in familiarity (global memory strength),[5] which is consistent with the descriptive analysis of the data reported by Verde himself. On the basis of the dual-process view, the difference in results between the present study and Verde's study thus suggests that associative- and item-recognition tasks lead to qualitatively different effects on unpracticed items' recollection and familiarity.

When fitting the single-process signal detection model to Verde's (2004) data, the restrictive equal-variance signal-detection model satisfactorily fitted the data. The fit was hardly improved when applying the unequal-variance variant of the single-process model. The result from the single-process analysis is consistent with the proposal that retrieval practice in Verde's experiment reduced the general memory strength of the unpracticed material. This single-process interpretation of the data is more parsimonious than the dual-process interpretation. In particular, it suggests a simple solution to the ostensible discrepancy between Verde's and our findings. To resolve this discrepancy, the dual-process view has to assume that retrieval practice has qualitatively different effects in associative- and item-recognition tasks. In contrast, the single-process view can account for the results in the two types of tasks on a purely quantitative basis, that is, by assuming that the amount of reduction in unpracticed items' general memory strength can vary across type of task.

Conclusions

Following dual-process accounts of recognition memory, the present results suggest that retrieval practice reduces unpracticed items' familiarity while leaving their recollection largely unaffected. These results disagree with previous evidence for reduced recollection in unpracticed items' free and cued recall and unpracticed items' associative recognition and thus challenge the dual-process account. Following the single-process account of recognition memory, the present results suggest that retrieval practice does nothing else than reduce unpracticed items' general memory strength. This indication is supported by the excellent fit of the model to the data. In particular, it agrees with the inhibitory account of retrieval-induced forgetting and is consistent with the results from prior work that examined retrieval-induced forgetting in free recall, cued recall, and response latency tasks. We therefore conclude that retrieval-induced forgetting in item recognition is caused by a reduction in the unpracticed items' general memory strength.

[5] Applying the independence remember–know model to the aggregate data reported in Verde's (2004) Experiment 1 yielded a larger decrease in recollection (.16 vs .22) than in familiarity (.09 vs .12) for unpracticed pairs compared with control pairs. Similarly, applying the STREAK model led to a larger decrease in specific memory strength (d_y: .55 vs .71) than in global memory strength (d_x: .21 vs .31); neither of the two effects was reliable, however, with both exhibiting the following: $\chi^2(1) < 2.90$, $p > .05$. The equal-variance signal-detection model satisfactorily fitted the data for unpracticed pairs, $\chi^2(1) = 2.12$, $p > .10$, and for control pairs, $\chi^2(1) = 2.38$, $p > .10$. The parameter estimates suggest significantly lower general memory strength for unpracticed pairs compared with control pairs (d': 0.56 vs. 0.74), $\chi^2(1) = 6.40$, $p < .05$. Applying the saturated unequal-variance signal-detection model yielded no variation of the strength distribution variance (σ: 1.25 vs. 1.23), $\chi^2(1) < 1.00$, but did show a reliable decrease in d' (0.50 vs. 0.71), $\chi^2(1) = 4.95$, $p < .05$.

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

Aslan, A., Bäuml, K.-H., & Pastötter, B. (2007). No inhibitory deficit in older adults' episodic memory. *Psychological Science, 18,* 72–78.

Battig, W. F., & Montague, W. E. (1969). Category norms of verbal items in 56 categories: A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology, 80,* 1–64.

Bäuml, K.-H. (2002). Semantic recall can cause episodic forgetting. *Psychological Science, 13,* 356–360.

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

Bäuml, K.-H., & Kuhbandner, C. (2007). Remembering can cause forgetting—but not in negative moods. *Psychological Science, 18,* 111–115.

Bäuml, K.-H., Zellner, M., & Vilimek, R. (2005). When remembering causes forgetting: Retrieval-induced forgetting as recovery failure. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31,* 1221–1234.

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

Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology, 58,* 17–22.

Dunn, J. C. (2004). Remember/know: A matter of confidence. *Psychological Review, 111,* 524–542.

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

Gómez-Ariza, C. J., Lechuga, M. T., Pelegrina, S., & Bajo, M. T. (2005). Retrieval-induced forgetting in recall and recognition of thematically related and unrelated sentences. *Memory & Cognition, 33,* 1431–1441.

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

Hockley, W. E., & Consoli, A. (1999). Familiarity and recollection in item and associative recognition. *Memory & Cognition, 27,* 657–664.

Humphreys, M. S. (1978). Item and relational information: A case for context independent retrieval. *Journal of Verbal Learning and Behavior, 17,* 175–187.

Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases.* Cambridge, MA: Cambridge University Press.

Macrae, C. N., & MacLeod, M. D. (1999). On recollections lost: When practice makes imperfect. *Journal of Personality and Social Psychology, 77,* 463–473.

Malmberg, K. J., & Xu, J. (2006). The influence of averaging and noisy decision strategies on the recognition memory ROC. *Psychonomic Bulletin & Review, 13,* 99–105.

Mannhaupt, H.-R. (1983). Reproduktionsnormen für verbale Reaktionen zu 40 geläufigen Kategorien [Production norms for verbal reactions to 40 familiar categories]. *Sprache und Kognition, 4,* 264–278.

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.

Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 803–814.

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

Rotello, C. M., Macmillan, N. A., & Reeder, J. A. (2004). Sum–difference theory of remembering and knowing: A two-dimensional signal detection model. *Psychological Review, 111,* 588–616.

Rotello, C. M., Macmillan, N. A., Reeder, J. A., & Wong, M. (2005). The remember response: Subject to bias, graded, and not a process-pure indicator of recollection. *Psychonomic Bulletin & Review, 12,* 865–873.

Rotello, C. M., Macmillan, N. A., & Van Tassel, G. (2000). Recall-to-reject in recognition: Evidence from ROC curves. *Journal of Memory and Language, 43,* 67–88.

Saunders, J., & MacLeod, M. D. (2002). The role of retrieval-induced forgetting in misinformation effects. *Journal of Experimental Psychology: Applied, 8,* 127–142.

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

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

Stretch, V., & Wixted, J. T. (1998). On the difference between strength-based and frequency-based mirror effects in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24,* 1379–1396.

Tulving, E. (1985). Memory and consciousness. *Canadian Psychology, 26,* 1–12.

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.

Verde, M. F. (2004). The retrieval practice effect in associative recognition. *Memory & Cognition, 32,* 1265–1272.

Verde, M. F., & Rotello, C. M. (2004). ROC curves show that the revelation effect is not a single phenomenon. *Psychonomic Bulletin & Review, 11,* 560–566.

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

Wixted, J. T. (2007). Dual-process theory and signal-detection theory of recognition memory. *Psychological Review, 114,* 152–176.

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

Wixted, J. T., & Stretch, V. (2004). In defense of the signal detection interpretation of remember/know judgements. *Psychonomic Bulletin & Review, 11,* 616–641.

Yonelinas, A. P. (1994). Receiver operating characteristics in recognition memory: Evidence for a dual-process model. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 1341–1354.

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language, 46,* 441–517.

Yonelinas, A. P., & Jacoby, L. L. (1995). The relation between remembering and knowing as bases for recognition: Effects of size congruency. *Journal of Memory and Language, 34,* 622–643.

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

(*Appendix follows*)

Appendix

Single-Process Signal-Detection Models

According to the single-process signal-detection model, recognition memory relies on a single source of memorial information, and recognition performance can thus be entirely described by detection theory: Subjects respond "old" whenever their assessment of an item's memory strength exceeds a given response criterion c_i (see Figure A1). Assuming equal variances of the underlying strength distributions, the probability of correctly recognizing a studied item (H) is given by

$$p(\mathrm{H}) = 1 - \Phi(c_i - d'),$$

and the probability of incorrectly recognizing a new item (FA) is given by

$$p(\mathrm{FA}) = 1 - \Phi(c_i),$$

where d' is the distance between the means of the underlying distributions for new versus old items. Across different levels of response confidence (c_i), the equal-variance signal-detection model theoretically yields a curvilinear receiver operating characteristic (ROC) that is symmetrical along the diagonal (see Panel C in Figure A1).

ROC analysis of recognition performance across varying levels of response confidence often yields ROC curves that are asymmetrical along the diagonal. According to detection theory, such asymmetrical ROCs indicate that the variance of the target distribution exceeds the variance of the noise distribution (see Panel B in Figure A1). Therefore, the model's equal-variance restriction is usually relaxed by letting the variance of the target distribution vary freely, which leads to

$$p(\mathrm{H}) = 1 - \Phi[(c_i - d')/\sigma],$$

where σ is the variance of the old items' distribution given that the variance of the new items' distribution is set to 1. Panel D in Figure A1 shows a typical theoretical ROC generated by the unequal-variance signal-detection model.

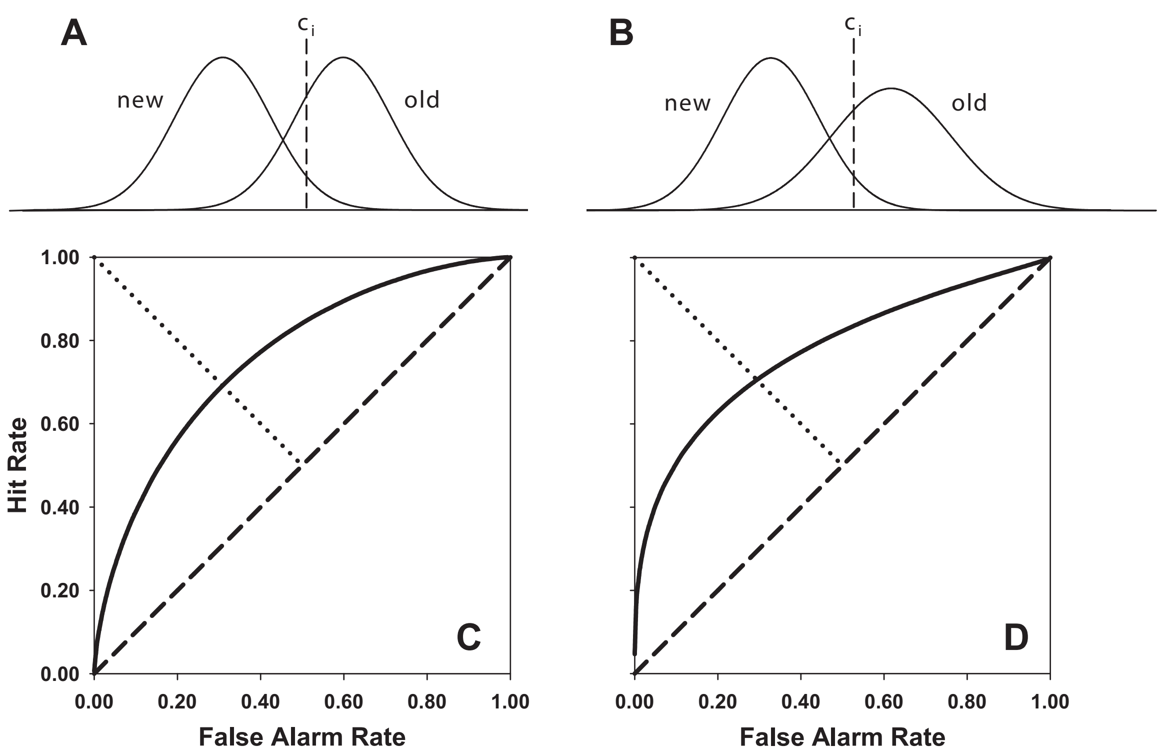

Figure A1. (A) Solid lines: Schematically depicted strength distributions for studied (old) and unstudied (new) material as assumed by the equal-variance signal-detection model. Dashed line: Location of response criterion. (B) Solid lines: Schematically depicted strength distributions for studied (old) and unstudied (new) material as assumed by the unequal-variance signal detection model. Dashed line: Location of response criterion. (C) Solid line: Theoretical receiver operating characteristics (ROC) predicted by the equal-variance model. Dashed line: Random guessing ROC. (D) Solid line: Theoretical ROC predicted by the unequal-variance model. Dashed line: Random guessing ROC.

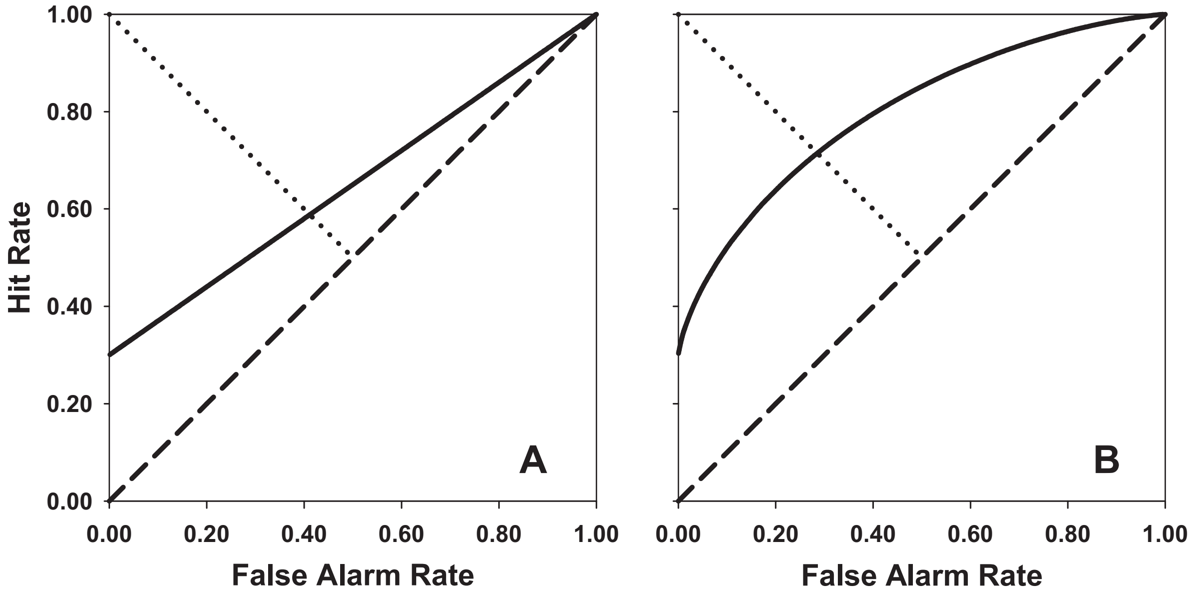

Figure A2. (A) Solid line: Theoretical receiver operating characteristics (ROC) predicted by the dual-process model if recognition is exclusively based on recollection. Dashed line: Random guessing ROC. (B) Solid line: Theoretical ROC predicted by the dual-process model if recognition is based on both recollection and stimulus familiarity. Dashed line: Random guessing ROC.

The Dual-Process Signal-Detection Model

According to the dual-process signal-detection model (Yonelinas, 1994), the characteristic shape of recognition ROCs results from the coaction of two qualitatively different memory processes—recollection and familiarity—which independently contribute to recognition performance. Considered in isolation, recollection is characterized as a threshold process that theoretically produces a linear ROC (see Panel A in Figure A2), whereas familiarity can be described by a classical signal-detection process as outlined above for the equal-variance single-process model (see Panels A and C in Figure A1). Formalization of these processes allows for decomposition of the asymmetrical item recognition ROC into a probability of recollection (R) and a signal-detection process with sensitivity d' corresponding to familiarity.

The model assumes that hits (H) can result from recollection or, in the absence of recollection, if an old item's familiarity exceeds the response criterion (c_i):

$$p(\mathrm{H}) = R + (1 - R)[1 - \Phi(\mathrm{c}_i - d')].$$

The probability for false alarms (FA) is given by the probability that a new item's familiarity exceeds the response criterion:

$$p(\mathrm{FA}) = 1 - \Phi(c_i),$$

because it is usually assumed that new items cannot be recollected. Panel B in Figure A2 shows a typical theoretical ROC generated by the dual-process model.

Parameter Estimation and Statistical Testing

Assuming binomially distributed response probabilities, the parameters of a recognition model can be estimated by minimizing the function

$$G^2 = \sum_i \left[2Hits_i \log\frac{\hat{p}(H)_i}{p(H)_i} + 2(n - Hits_i)log\frac{1 - \hat{p}(H)_i}{1 - p(H)_i} \right]$$

$$+ \sum_i \left[2FAs_i \log\frac{\hat{p}(FA)_i}{p(FA)_i} + 2(n - FAs_i)log\frac{1 - \hat{p}(FA)_i}{1 - p(FA)_i} \right]$$

where $\hat{p}(H)_i$ and $\hat{p}(FA)_i$ are the observed hit and false alarm rates at confidence level i, and $p(H)_i$ and $p(FA)_i$ are the model equations predicting these rates as functions of the model's parameters. $Hits_i$ and FAs_i reflect the absolute frequencies of observed hits and false alarms at confidence level i. The minimized G^2 function yields maximum-likelihood estimates for the model's parameters plus a χ^2-distributed G^2 statistic for testing the model's goodness of fit. The maximum-likelihood procedure can be analogously applied to the signal-detection models outlined above as well as to the sum–difference theory of remembering and knowing model.

Received October 26, 2006
Revision received April 23, 2007
Accepted April 26, 2007 ■