Retrieval-Induced Forgetting and Recognition Memory: Insights from Behavioral and Electrophysiological Experiments

Inaugural-Dissertation zur Erlangung der Doktorwürde
der Philosophischen Fakultät II
(Psychologie, Pädagogik und Sportwissenschaft)
der Universität Regensburg

vorgelegt von

BERNHARD JOHANNES SPITZER

aus Neustadt a.d. Waldnaab

Regensburg 2007

Erstgutachter: Prof. Dr. Karl-Heinz Bäuml
Zweitgutachter: Prof. Dr. Mark W. Greenlee

Acknowledgement

The present thesis had never been written without the scientific influence, support, and inspiration of Prof. Dr. Karl-Heinz Bäuml. He awoke my interest in memory research and provided me with the infrastructure for the reported experiments, while always being a proficient and helpful teacher.

Since the first day of my work, I owe very much to Dr. Alp Aslan, for thousands of shared coffee breaks, fruitful discussions and memorable conferences abroad. Many thanks to Dr. Simon Hanslmayr for his tuition in matlab programming and for expanding my vocabulary. Without exception, the colleagues in our constantly growing team not only enriched my research activities but also were a pleasure to work and to non-work with. In particular, I want to thank Maria Wimber, Bernhard Pastötter, Christof Kubandner, Claus Arnold, and Dr. Martina Zellner, for always lending me their helping hands, and for lots of fun in- and outside the university.

Many thanks to Verena Bauer, Matthias Gruber, Philipp Leipold, Klaus Nuissl, Renate Schmidtner, and Markus Wildenauer, as well as to many nameless students for their help with gathering the present data.

In a completely different sense, I want to thank one certain little smurf for cheering me up and keeping me going.

However, this work is dedicated to my parents, for their loving support throughout all episodes of my life leading me to now and further.

Preface

One of the most intriguing features of human memory is its ability to forget. In everyday life, the forgetting of certain episodes, knowledge, or valuable information can be embarrassing, annoying, or expensive. On the other hand, getting rid of irrelevant, unpleasant, or interfering memories can make our existence a good deal easier to bear. Psychological research has identified and investigated a large variety of memory phenomena which can contribute to our daily experience of forgetting. Within this research field, a particularly paradoxical finding has emerged in recent years: Remembering certain information can cause forgetting of related information (*Retrieval-induced forgetting*, Anderson, Bjork, & Bjork 1994). Experimental research on the topic has yielded strong evidence that this type of forgetting is mediated by a distinct memory mechanism: Inhibition. One basic goal of the present work was to investigate the generality of retrieval-induced forgetting in novel experimental settings. In five experiments, such forgetting could be reliably demonstrated across a variety of experimental setups, supporting the view of inhibition as a general and highly relevant memory mechanism.

Both in everyday life and in memory research we are not only concerned with the question why or how we forgot something, we also ask ourselves: What happened to the things we once experienced or knew? Have these memories disappeared? Have they faded and become too pale to discover? Are we just unable to find them although they are still perfectly intact somewhere in our heads? Are they repressed in a Freudian sense? The present study is concerned with a similar question regarding retrieval-induced forgetting: What are the concrete effects of this type of forgetting on the affected memories? How has the memory representation of "forgotten" information changed, i.e. quantitatively, qualitatively, or both? What has inhibition done to these memories?

Forgetting can manifest itself in various ways: We forget what exactly we did 43 days and 8 hours ago, we forget passwords, and sometimes we also fail to recognize somebody we have had a charming conversation with only days before. In experimental memory research, recognition tests have proven extremely useful to characterize memory phenomena and their effects on memories: It is much easier for subjects to recognize a presented item than e.g. to recall it freely, and recognition tests may thus yield information about memories that would have been 'forgotten' in many other memory testing procedures. According to many prominent theories of recognition memory, a feeling of 'familiarity' with a presented stimulus may be sufficient to correctly judge it as "seen before", even if the subject can not retrieve any particular information associated with the prior occurrence of that stimulus. Numerous forgetting phenomena affecting episodic recall are known to have no detrimental effects on recognition memory and it has thus been argued that these types of forgetting selectively disrupt retrieval processes. Retrieval-induced forgetting, however, has been recently demonstrated with recognition testing procedures as well (e.g. Hicks & Starns, 2004, Verde, 2004). Based on these notable empirical findings, the present study is designed to investigate the effects of retrieval-induced forgetting on recognition memory in more detail.

Technically, standard recognition tests can easily be extended for obtaining a more complex data structure that characterizes performance both quantitatively and qualitatively (see section 2). Various mathematical models have been proposed to interpret such data matrizes as a function of cognitive parameters. For the present empirical data, I used a variety of modeling approaches to give a detailed description of the quantitative and qualitative effects of retrieval-induced forgetting in recognition memory. However, there is still considerable disagreement in the literature on how to characterize recognition performance appropriately. Although it is beyond the scope of the present work to resolve this theoretical dispute, the present experiments yielded a large amount of data allowing for an informative comparison of opposing models both on technical and theoretical grounds. In doing so, the present work not only provides the reader with novel empirical findings in the light of different analysis methods, it also suggests that the effects of retrieval-induced forgetting can be fully accommodated by a surprisingly simple theoretical model of recognition memory.

Finally, recognition tests are invaluable in neuropsychological memory research, because they allow for investigating the (e.g., electro-)physiological correlates of memory processes triggered by the controlled presentation of studied vs unstudied stimuli. In the present study, using advanced EEG methodology, an electrophysiological correlate of recognition memory was identified that appears to be specifically sensitive to the inhibitory mechanisms underlying retrieval-induced forgetting. To date, these electrophysiological data may mainly be of descriptive value, given the considerable theoretical discrepancies between prominent behavioral and neuropsychological models of recognition memory. Irrespective of these, however, advanced analysis of the present EEG data further revealed novel recognition correlates, which might contribute to a better understanding of the complex neural mechanisms underlying recognition memory.

Contents

Parts of the present thesis have been accepted for publication in *Journal of Experimental Psychology: Learning, Memory, & Cognition*:
Spitzer, B. & Bäuml, K.-H. (in press). Retrieval-induced forgetting in item recognition: evidence for a reduction in general memory strength.

Abstract

Retrieving a subset of previously studied items can impair later recognition of related items. Using the remember/know procedure (Experiment 1 and 5), the ROC procedure (Experiments 2-4), and analysis of EEG old/new effects (Experiment 5), it was investigated how such retrieval-induced forgetting can be explained in terms of single-process and dual-process accounts of recognition memory. Across experiments, dual-process analysis yielded a complex pattern of results which suggests that retrieval practice can affect familiarity and - less reliably - also recollection of the unpracticed material. 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 suggests that retrieval practice reduced unpracticed items' general memory strength in each of the experiments. 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. Recognition of unpracticed material was further characterized by reduced theta power in the EEG, indicating that unpracticed material triggered only relatively weak memory signals. It is argued that the detrimental effect of retrieval practice is best characterized as a weakening of the unpracticed item's inherent memory representation. Analysis of inter-electrode synchronisation further revealed a novel type of recognition correlate in the theta frequency range.

Part I

Background

Retrieval-Induced Forgetting

Retrieving a subset of formerly studied material can cause subsequent forgetting of the nonretrieved material. Such retrieval-induced forgetting has repeatedly been demonstrated using 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*). 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 using 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 (*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, i.e., the forgetting of unpracticed items is only observed when practice on a subset of the studied exemplars requires active retrieval of the material, whereas merely strengthening these items by repeated exposure is not sufficient to cause forgetting (Anderson, Bjork, & Bjork, 2000; Bäuml, 2002; Ciranni & Shimamura, 1999). The forgetting effect 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 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 which employ so-called independent probes, i.e., probes which 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).

Many of the studies reported above investigated retrieval-induced forgetting using verbal material that could be grouped by specific semantic features. However, retrieval-induced forgetting has also been observed with visual material that was purely episodically related (Ciranni & Shimamura, 1999). In their experiments, retrieval practice on a subset of previously studied non-verbal objects that could be grouped by perceptual features caused forgetting of episodic information associated with perceptually related objects. Like with semantically categorized material (see above), retrieval-induced forgetting of episodically related material appeared to be retrieval-specific, suggesting that the forgetting may be mediated by the same inhibitory mechanisms in both experimental situations.

Recognition Memory

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). On the one hand, individuals can base their recognition judgments on conscious recollection of information about the spatio-temporal 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. While recollection is often conceptualized as a relatively slow process yielding qualitative information about previous events, familiarity reflects a relatively fast, strength-like 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 indices recollection ("remember" = R), while the proportion of know responses is thought to reflect familiarity in the absence of recollection ("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.

Using the ROC procedure, 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. 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 strength-like signal detection process that yields a symmetrical ROC, recollection is thought to be a threshold-like 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 Appendix).

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, i.e., 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 do not incorporate the independent contribution of a threshold-like recollection process to recognition performance. Rather, to account for the characteristic shapes of recognition ROCs, it is often assumed that the variance of the strength distribution for studied items can exceed the variance of the distribution for unstudied items (see Appendix for details). Such increased variability in the target items' strength can be explained by assuming that the single items on a list might be differentially strengthened during study exposure. When applied to recognition data, the standard signal detection model's equal-variance restriction is thus usually relaxed by letting the variance of the target variance vary freely. Extensive review and reanalysis of a large body of recognition data has shown that single-process signal detection models often lead to a more adequate description of empirical remember/know and ROC data than dual-process accounts (e.g., Dunn, 2004; Wixted, 2007; Wixted & Stretch, 2004).

It should be noted that signal detection models of recognition memory, although typically referred to as 'single-process' models, do theoretically not exclude the possibility that the psychological variable 'general memory strength' may be potentially produced by a variety of underlying processes (e.g. familiarity and recollection). It is rather argued that the dual-process models' particular conceptualization of recollection as an independent all-or-none process is flawed and that thus dual-process interpretations of remember/know and ROC data rely upon incorrect psychological assumptions (e.g., Dunn, 2004; Wixted, 2007; Wixted & Stretch, 2004). In single-process theory, both recollection and familiarity are treated as continuous variables that vary along a common dimension of general memory strength. Thus, the relative contribution of recollection or familiarity to recognition performance can not be disentangled on the basis of behavioral remember/know or ROC data. Of course, because a single strength parameter has proven sufficient to account for such data (e.g., Dunn, 2004; Wixted, 2007; Wixted & Stretch, 2004), the assumption of eventual subsidiary processes appears redundant. However, the possibility that such subsidiary processes (i.e., recollection, familiarity, or any other process) may exist is not precluded in single-process acounts of recognition memory.

Electrophysiological Correlates of Recognition Memory

While current behavioral memory research increasingly favours an alternative interpretation of recognition data, dual-process theory as proposed e.g. by Yonelinas (2002) is widely accepted in neuropsychological research on recognition memory (for reviews, see Curran et al., 2006 or Rugg & Yonelinas, 2003). A particularly prominent finding within this line of research has emerged from the analysis of event-related potentials (ERPs). The ERP old/new effect refers to the finding that correct recognition of old (i.e., formerly studied) items elicits different ERP waveforms at frontal and parietal recording sites compared to correct rejection of new (i.e., unstudied) items (e.g., Rugg & Doyle, 1992; Rugg et al., 1998).

In studies applying the remember/know procedure, early frontal (FN400) old/new effects have been reported both for 'known' and for 'remembered' items, whereas late (approx. 500-800 ms) parietal old/new effects are usually selectively enhanced for remember items (e.g., Curran, 2004; Woodruff et al., 2006). Further, dissociations of ERP old/new effects have been obtained across different recognition tasks that are believed to be predominantly recollection- or familiarity-driven (e.g., Curran, 2000; Curran & Cleary, 2003; Jäger, et al., 2006; Rugg et al., 1998). The common interpretation of such findings implies that the early frontal ERP old/new effect reflects stimulus familiarity, whereas the late parietal component is indicative of recollective processes (for reviews, see Curran et al., 2006; Friedman & Johnson, 2000 or

Rugg & Yonelinas, 2003).

However, the analysis of ERP old/new effects gives only an incomplete description of EEG activity associated with recognition memory. ERPs are formed by averaging time-locked EEG epochs across trials and thus retain only evoked brain activity that is phase-locked with respect to stimulus onset. Alternative EEG analysis methods have been proposed that allow for examination of non-phase locked (i.e., induced) oscillatory brain activity, e.g. in terms of event-related changes in band power due to (de-)synchronisation of neural firing (Pfurtscheller & Aranibar, 1977; Pfurtscheller, 1992).

With respect to recognition memory, it has been shown that event-related (de-)synchronisation (ERD/ERS, see method section of Experiment 5 for technical details) in the theta frequency range (approx. 4-7 Hz) exhibits early and late old/new effects that are topographically and temporally similar to those typically found with ERPs (Klimesch et al., 2000; 2001; 2006). Although it has been argued that the old/new effects in ERP and ERD/ERS measures are not directly related (e.g. Jacobs et al., 2006), early and late old/new effects in theta ERD/ERS have similarly been associated with familiarity and recollection (Klimesch et al., 2001). More specifically, it has been suggested that the degree of theta ERS after presentation of studied items basically reflects the strength of these items' episodic memory traces, but that the timecourse of such theta ERS may depend on the memory processes involved in evaluation of these memory traces: For familiarity-based evaluation, such theta ERS is thought to occur relatively early (200-400 ms), whereas for recollection-based evaluation, theta ERS appears to be significantly prolonged ($>$ 500 ms) (Klimesch et al., 2001; 2006). Thus, the dual-process account of theta ERS old/new effects (Klimesch et al., 2001; 2006) theoretically suggests that the two processes may be contingent upon a common underlying dimension of memory strength.

A large majority of EEG and fMRI studies examining correlates of recognition memory has relied on dual-process theory, but the topic remains controversial. For example, alternative interpretations for ERP old/new effects have been proposed (e.g., Finnigan et al., 2002; Voss & Paller, 2006; Yovel & Paller, 2004). Further, recent attempts to identify brain regions that are associated with recollection and/or familiarity yielded inconsistent or even conflictive outcomes across various fMRI studies (e.g., Gonsalves et al., 2005; Manns et al., 2003; Vilberg & Rugg, 2007; Wais et al., 2006; Yonelinas et al., 2005).

However, beyond such partial disagreement with respect to the very neural substrates of recollection and familiarity, Wixted (2007) recently raised more general concerns regarding the common acceptance of dual-process theory in neuropsychology. As outlined in the previous section, the dual-process models' theoretical conceptualization of recollection and familiarity has been heavily criticised on the basis of a large body of behavioral evidence. Thus, the majority of neuropsychological recognition studies might have employed experimental designs and/or theoretical interpretations that rely on a possibly incorrect psychological model (Wixted, 2007). If true, of course, the specific functional relevance of the electrophysiological recognition correlates outlined above may be open to alternative interpretations in the future.

Retrieval-Induced Forgetting and Dual-Process Models

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 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 Models

Single-process signal detection models often lead to a more adequate description of empirical remember/know and ROC data than 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.

Retrieval-Induced Forgetting of Episodically Related Material

The phenomenon of retrieval-induced forgetting is not restricted to the study and retrieval of semantically related words, but has also been demonstrated with purely episodically related material (e.g. Ciranni & Shimamura, 1999). In their experiments, geometrical objects that could be grouped either by color or by shape were presented in different locations for study. Subjects were subsequently asked to recall specific features (e.g. color, shape, or location) of a subset of the stimuli. In a final test comprising all previously studied objects, recall of these features was impaired for unpracticed stimuli that shared perceptual features (e.g. shape) with the practiced material. To date, however, it remains unclear to what extent such retrieval-induced forgetting of episodically related material affects recognition memory as well.

In the light of dual-process models of recognition memory, it may well be that the effects of retrieval inhibition on episodically as compared to semantically related memories are qualitatively different. With purely episodically related (i.e., pre-experimentally unrelated) material, interference, inhibition and subsequent forgetting of unpracticed material may largely depend upon contextual (i.e., recollective) information acquired during the study episode (see Anderson, 2003, for related considerations). Thus, if retrieval inhibition of unpracticed material serves to reduce interference of information competing during retrieval practice, it can be hypothesized that retrieval-induced forgetting of purely episodically related material may primarily lower the availability of recollective information rather than affecting stimulus familiarity.

However, the results from Ciranni and Shimamura's (1999) experiments suggest that the detrimental effects of retrieval-induced forgetting on episodically related material were not restricted to specific item features, but rather appeared to affect the general memory representation of the unpracticed material. In particular, if a certain feature (e.g. location) was retrieval practiced, not only recall of that feature (location) but also recall of other features (e.g. shape) was impaired for unpracticed objects. Such generalized forgetting may indeed reflect a decrease in the unpracticed material's general memory strength rather than merely a reduction in episodic recollection. As outlined in the previous section, the effects of retrieval practice on semantically related material may be described best by a single-process model of recognition memory. Suggested from Ciranni and Shimamura's (1999) evidence for a reduction in general memory strength, the same may hold for retrieval-induced forgetting of episodically related material.

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). One major 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. Five experiments are reported, in each of which participants studied categorized material. After retrieval practice on a subset of the material, recognition tests were employed. To investigate possible qualitative effects of retrieval practice on recognition memory, the remember/know procedure was applied in Experiments 1 and 5, and the ROC procedure was applied in Experiment 2-4. Further, in Experiment 5, possible EEG correlates of retrieval-induced forgetting in item recognition were investigated. Using a large variety of prominent dual-process measurement methods, the present study was designed to provide an answer on whether retrieval practice impairs only recollection of unpracticed material or affects its familiarity as well.

The second major 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. The design of the present study allows for a detailed comparison of dual- and single-process models within and across experiments, on the basis of both behavioral and electrophysiological data. 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).

Another motivation for the present work was to investigate the generality and/or possible limitations of the effects of retrieval-induced forgetting on recognition memory. Therefore, variations of the standard retrieval-practice paradigm were designed: Whereas in Experiments 1, 2, and 5, semantically categorized material was used to induce forgetting, the study material used in Experiments 3 and 4 was purely episodically structured. Of particular interest was the question whether retrieval interference - and thus forgetting - can arise on the basis of shared perceptual contexts (Experiment 3) and/or on the basis of shared perceptual features (Experiment 4) of the study material. It was further examined under which of these conditions retrieval-induced forgetting potentially affects not only item recognition performance but also memory for episodic information associated with the study material. The results from these experiments may help identifying factors that possibly generalize or moderate the effects of retrieval inhibition on related memories.

Finally, the present work is concerned with a remarkable conceptual discrepancy in current recognition memory research: While highly disputed on the basis of behavioral evidence, current neuropsychological research on recognition memory is almost exclusively based on dual-process theory. From analysis of Experiment 5, it can be determined whether the effects of retrieval practice on a large variety of behavioral dual-process measures are compatible with the effects on putative EEG correlates of recollection and familiarity, thus allowing for a test of the neurocognitive dual-process framework's internal consistency. Further, the data from Experiment 5 can be used for a detailed analysis of the electrophysiological correlates of remembering and knowing. Advanced methodology was employed to examine not only 'evoked' brain activity but also 'induced' activity associated with subjective reports of recollective experiences. In combination with the results from the behavioral experiments, the detailed analysis of Experiment 5 may potentially help refining the theoretical conceptualization of episodic recollection, which currently is the main point of disagreement between opposing models of recognition memory.

Part II

Retrieval-Induced Forgetting and Recognition Performance

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, a decrease in remember responses was expected 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

52 adults (18-40 years old) took part in the experiment on a voluntary basis. They were tested individually in sessions that lasted approximately 30 minutes.

Material

Prior work reported retrieval-induced forgetting in DRM "categories" by showing that retrieval practice on half of the noncritical items can cause forgetting of the remaining noncritical items (Bäuml & Kuhbandner, 2003). In the present experiment, stimuli consisted of 12 exemplars from each of eight target categories selected from the 24 DRM-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 (rank 4-9) were studied together with their critical item serving as category cue. The three strongest (rank 1-3) and the three weakest associates (rank 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 (rank 7-9) while the three stronger associates (rank 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). 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, it was 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 sec per item. The serial order of the items was block randomized, i.e., 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, the remaining three were shown at the end. After half of the participants, the order of the items was reversed. After the study phase, participants were instructed to count backwards from 300 in steps of threes for 60 sec 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 sec 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, the order was reversed. Following the retrieval-practice phase, a questionnaire was presented in which participants were asked to rate German politicians on dimensions like 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 out 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 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 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 Performance

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+), the remember hit rate was .51 for practiced items and .34 for their controls. A 2 × 2 analysis of variance with the two factors of item type (practiced vs. control) and criterion (old/new vs. remember) showed a significant main effect of item type [$F(1, 51) = 65.66$, $MSE = 0.026$, $p <$

Table 1: Hit rates and false alarm rates in Experiment 1. Upper panel: Old/new recognition performance. Lower Panel: Remember performance.

	Hits		False Alarms	
Item Type	Rate	*SE*	Rate	*SE*
		Recognition		
P+	0.85*	0.02	0.22	0.02
C+	0.65	0.03	0.20	0.02
P-	0.60*	0.03	0.22	0.02
C-	0.68	0.02	0.20	0.02
		Remember		
P+	0.51*	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: Asterisks indicate significant deviations from control performance ($p < .01$). P+ = practiced items; P- = unpracticed items; C+ = control items for the practiced items; C- = control items for the unpracticed items. SE = standard error.

.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-), the remember hit rate was .34 for unpracticed items and .37 for their controls. A 2 × 2 analysis of variance 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 [$F(1, 51) = 144.04$, $MSE = 0.031$, $p < .001$, 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 [$t(51) = 1.20$, $p > .20$, two-tailed] but significantly fewer recognition hits for unpracticed material [$t(51) = 2.85$, $p < .01$, $d = .43$, two-tailed].

False alarm rates were examined separately. As expected, a 2 × 2 analysis of variance 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 [$F(1, 51) = 1.64$, $MSE = 0.002$, $p > .30$] was 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 compliment 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.[1]

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 analysis of variance 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 [$t(51) = 4.47$, $p < .001$, $d = .61$, two-tailed] and familiarity [$t(51) = 4.10$, $p < .001$, $d = .49$, two-tailed] of the practiced items.

[1] 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 were 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.

The recollection estimate for unpracticed items (P-) was .27 compared to .31 for their controls (C-). The familiarity estimates were .23 for P- items and .31 for C- items. A 2×2 analysis of variance 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 [$t(51) = 2.02$, $p < .05$, $d = .34$, two-tailed] but no reliable reduction in recollective processes [$t(51) = 1.31$, $p > .15$, two-tailed].

Also, the STREAK model (Rotello, Macmillan, & Reeder, 2004) was fitted 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). Following Rotello et al. (2004), parameter s was set to 0.8 for each item type.

The best fitting STREAK parameters for both global (d_x) and specific (d_y) 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 were 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 can

Table 2: Dual-process parameter estimates for Experiment 1. Upper panel: Independence remember/know model. Lower Panel: STREAK model.

	Independence Remember/Know			
Item Type	F	R		
P+	.45*	.44*		
C+	.29	.28		
P-	.23*	.27		
C-	.31	.31		
	STREAK			
Item Type	d_x	d_y	C_0	C_r
P+	1.00*	1.46*	0.61	0.24
C+	0.64	0.97	0.67	0.06
P-	0.50*	0.90	0.61	0.14
C-	0.68	1.05	0.67	0.10

Note: Asterisks indicate significant deviations from control performance ($p < .05$). P+ = practiced items; P- = unpracticed items; C+ = control items for the practiced items; C- = control items for the unpracticed items. F denotes familiarity, R recollection. d_x denotes global memory strength, d_y specific memory strength, C_0 location of the STREAK old/new criterion, and C_r location of the STREAK remember criterion.

not be tested statistically. However, likelihood-ratio tests can be used to examine whether the model's parameters vary with item type. For practiced (P+) items, the increase in global as well as in 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 unpracticed items (P-), only global memory strength d_x was significantly reduced [$\chi^2(1)= 5.05$, $p < .05$], whereas the decrease in specific memory strength 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 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 practiced (P+) items [$\chi^2(1) = 8.39$, $p < .005$]. Therefore, the more general unequal-variance signal detection model was fitted to the data, allowing the variance of the old items' distribution (parameter σ) 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 can not 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 Table 3).

Table 3: Experiment 1, single-process parameter estimates derived from the unequal-variance signal detection model.

Item Type	d'	σ	k	r
P+	1.50*	0.71	0.76	1.49
C+	1.18	0.88	0.84	1.54
P-	1.04*	1.05	0.76	1.49
C-	1.25	0.87	0.84	1.54

Note: Asterisks indicate significant deviations from control performance ($p <$ .01). P+ = practiced items; P- = unpracticed items; C+ = control items for the practiced items; C- = control items for the unpracticed items. d' denotes general memory strength, σ variance of the target distribution, k location of the know criterion, and r location of the remember criterion.

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 practiced items (P+) compared to their controls (C+) (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 unpracticed items (P-) compared to their controls (C-) (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 σ [$\chi^2(3)$= 4.85, $p > .15$].[2] [3]

Discussion

Using a standard item recognition task, the results of Experiment 1 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

[2] 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 σ 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 σ 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.

[3] 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). I 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.

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 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 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, ROC methods were used 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 evaluation of the two types of models.

Experiment 2

Using 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. Therefore, the ROC procedure was used in Experiment 2 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 model's ability to account for the recognition data can be tested statistically. These data will help evaluating 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 minutes.

Material

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 (rank 4-15 in the norms). Within each of the eight categories, six of the chosen exemplars (rank 7-12) were studied, and the remaining six (rank 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 Experiment 1, with the exception that retrieval practice was not restricted to weaker exemplars. For four of the eight target categories, participants retrieval practiced three of the six studied exemplars, with each item of a category serving equally often as a practiced and unpracticed item across participants. 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, it was counterbalanced which categories were retrieval practiced.

Procedure

Participants were tested individually in a quiet surrounding, seated in front of a 15” 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 2500 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 they were instructed to solve as many problems as possible within 2 minutes.

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 participant's 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 of 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 participant's 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.

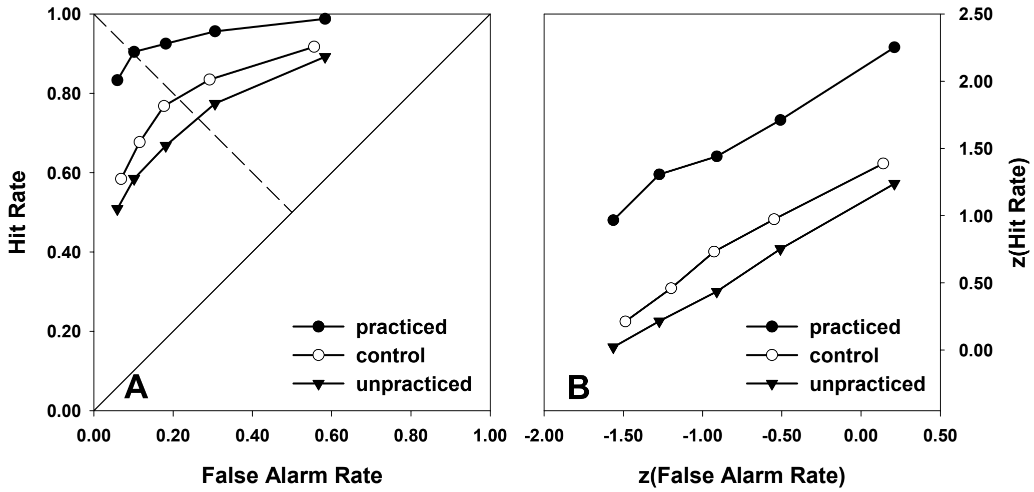

Figure 1: Recognition ROCs (A) and zROCs (B) for the three item types in Experiment 2.

Recognition Test

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

For dual-process analysis, the dual-process signal detection model (Yonelinas, 1994) was fitted to the raw data, assuming a threshold-like recollection process in addition to a strength-like signal detection familiarity process (for details, see 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 ten 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 using maximum likelihood techniques which also allow for statistical testing.

The dual-process model described the data for practiced (P+) and unpracticed items (P-) well [P+: $\chi^2(3) = 2.92$, $p > 0.40$; P-: $\chi^2(3) = 1.38$, $p > 0.70$], but could not satisfactorily fit the data for the control items (C) [$\chi^2(3) = 9.14$, $p < 0.05$]. Still, stable maximum likelihood parameter estimates could be derived for each item type. The parameter estimates for recollection (R) and familiarity (as measured by d') are depicted in Table 4. The estimates for d' were 1.65 for practiced items (P+), 0.80 for unpracticed items (P-), and 1.10 for control items (C). The estimates for R were 0.67 for practiced items (P+), 0.38 for unpracticed items (P-), and 0.40 for control items (C).

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 practiced items (P+) than for control items (C) 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 unpracticed items (P-) and control items (C) 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

Like in Experiment 1, the equal-variance signal detection model was fitted 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 5 response criteria) to fit the ten data points in an old/new rating experiment. The equal-variance model could not describe the recognition ROCs and had to be

Table 4: Dual-process and single-process parameter estimates for Experiment 2. Upper panel: Dual-process signal detection model. Lower panel: Unequal-variance signal detection model.

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

Note: Asterisks indicate significant deviations from control performance ($p < .05$). P+ = practiced items; P- = unpracticed items; C = control items. d' denotes familiarity (dual-process model) or general memory strength (unequal-variance model), R denotes recollection, and σ denotes variance of the target distribution. *df* denotes the model's degrees of freedom for statistical testing. ‡ indicates insufficient goodness of fit.

rejected for all three item types [all $\chi^2(4)$'s$>$ 14.00, all p's$<$ .01]. Therefore, the more general unequal-variance signal detection model was fitted to the data. When applied to 5-point ROC data, the unequal-variance signal detection model has seven free parameters (sensitivity d', target variance σ, and 5 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 $p's > .30$]. 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 practiced items (P+) than for control items (C) (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 unpracticed items (P-) than for control items (C) (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 σ nor the placement of the five confidence criteria varied reliably across item type ([$\chi^2(2) = 1.53, p > 0.40$], [$\chi^2(10) = 5.83, p > 0.80$]).

Discussion

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

Experiment 3

In cued recall, Ciranni and Shimamura (1999) found retrieval-induced forgetting with episodically related material that is comparable to the one usually observed in the standard retrieval-practice paradigm with semantically related material. To date, however, retrieval-induced forgetting of episodically related material has not been examined with recognition testing yet.

In Experiment 3, a variation of the retrieval-practice paradigm was designed. Subjects studied semantically unrelated words that were displayed together with different context colors. During retrieval practice, a subset of the items from half of these artificially created color categories was then repeatedly retrieved. In a subsequent item recognition test, the items from the study list had to be discriminated from new items. Again, the ROC-procedure was used to investigate the qualitative effects of retrieval practice in recognition memory. Following the item recognition test, subjects were additionally asked to remember each studied item's context color in order to explore whether the expected effects of retrieval-induced forgetting generalize to the recall of episodic features of the unpracticed material, as suggested by Ciranni and Shimamura (1999).

Following the dual-process model's conceptualization of recollection as an episodic, recall-like process (e.g. Yonelinas, 2002), the use of episodically related material might enhance the role of recollective processes in the present experiment. However, the dual-process framework as proposed e.g. by Yonelinas (2002) showed some technical and conceptual difficulties in accommodating the results from the previous experiments, whereas a single-process account of the data suffered from none of these difficulties. Using different material, Experiment 3 should yield further insights regarding the competing models' appropriateness to accommodate the patterns of results across experimental contexts. Suggested from prior evidence for generalized forgetting of episodically related material (Ciranni & Shimamura, 1999; see introduction), however, single-process analysis of Experiment 3 may yield comparable modulations in general memory strength as Experiments 1 and 2 in which semantically related material was used.

Method

Participants

Subjects were 64 young adults (19-31 years old) who were paid 6 Euro for participating in the experiment. They were tested individually in sessions that lasted approximately 25 minutes.

Material

The primary stimuli were two sets of 48 semantically unrelated German words chosen from the norms reported in Mannhaupt (1983). The two sets of items were pairwise related, i.e., each word from set one was semantically associated to one of the words from set two. None of the chosen items was pre-experimentally associated with any particular color.

Design

The Experiment consisted of four main phases: A study phase, a retrieval practice phase, a recognition test phase, and a color recall phase. Across subjects it was counterbalanced which of the two sets of items was studied. The 48 study words were arbitrarily divided into four sets of 12 items. For each subject, the font colors red, green, blue, and yellow were assigned to the four "categories" in the study phase. Color assignment was counterbalanced, so that each item was equally often assigned to every color across subjects. From each of two of the "color categories", six target items were used in the retrieval practice phase. Accordingly, 12 items (each assigned to one of two colors, e.g. red and blue) were retrieval practiced (P+), 12 items (each assigned to one of the same colors as P+) were unpracticed (P-), and 24 Items (each assigned to one of the two remaining colors, e.g. green and yellow) served as control items (C). Assignment of the items to retrieval practice was counterbalanced, so that every item was practiced equally often across subjects. All 48 studied item-color pairings were used both in the recognition phase and in the color recall phase. The recognition phase additionally incorporated the set of 48 unstudied words, which were each assigned to the same color as their respective studied associate.

Procedure

Subjects were tested individually in a quiet surrounding, comfortably seated in front of a 15" TFT Computer screen. They were instructed to memorize each item on the study list together with its context color and that later on memory would be tested for both the words and the respective colors. Every item was presented in a black Arial font for 1500 ms, horizontally framed by two colored bars, which were displayed in the item's assigned color. Each item-context presentation was followed by a 500 ms blank screen. The serial order of the items was block randomized, with the restriction that no item was displayed with the same color as its antecessor. Furthermore, three buffer items were added at the beginning and at the end of the list and the buffer items were also randomly assigned to one of the four types of color bars. Following the study phase, subjects had to count backwards from 300 in steps of threes for approximately 60 seconds.

Subsequently, the retrieval practice phase started. The word stems were framed by the same color bars as the corresponding items in the study phase. Subjects were instructed to use the given colors as cues to complete the word-stems from memory. The serial order of the word-stems was quasi-random, with alternating colors. The experimenter noted the subjects response on a prepared data sheet and the subjects could proceed to the next item by pressing a key. After retrieval practice, subjects had to solve a set of simple arithmetic problems for about four minutes and were allowed to use a pocket calculator in order to avoid frustration in the subject.

Following the distractor task, an item recognition test based on confidence ratings was applied. The item-context pairings were presented together with a schematically depicted 6-point rating scale ranging from *1: definitely old word* to *6: definitely new word* in the lower part of the screen. Subjects were instructed to enter their responses via the digits of the PC-Keyboard. As soon as any numerical response was entered, the next item was presented on the screen; responses were automatically recorded in a log file. The serial order of the test items was block randomized in the same manner as in the previous experiments

After completion of the recognition test, the color recall phase started. Each initially studied item was displayed on the screen in the same manner as in the study phase, except that the previously colored bars were always displayed in a light gray. Subjects were instructed to indicate the original color of the bars by pressing one of four correspondingly colored keys. As soon as any response was entered, a response feedback ("correct" or "incorrect") was displayed on the screen for 1000 ms, then the next item was presented. The serial order of the test items was block randomized in the same manner as in the previous experiments.

Results

Retrieval-Practice Phase

On average, the participants successfully completed 48.0% ($SE = 0.02$) of the word stems 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

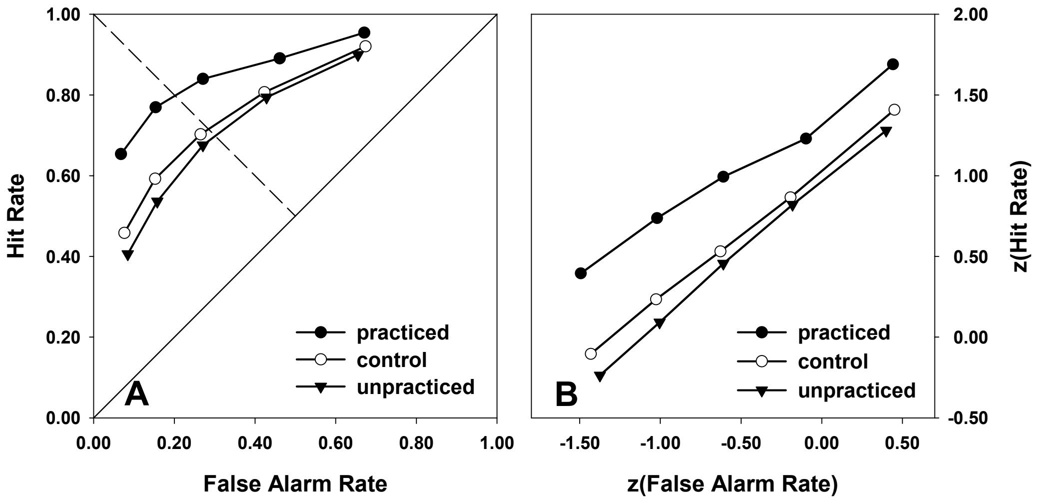

Figure 2: Item recognition ROCs (A) and zROCs (B) for the three item types in Experiment 3.

confident criterion point (*definitely old*). The ROCs obtained by plotting the cumulative false alarm against hit rates for each item type are illustrated in Figure 2.

Dual-Process Analysis

The dual-process model (Yonelinas, 1994) described the data set well [all $\chi^2(3)$'s$>$ 4.20, all p's$>$ 0.20] and stable maximum likelihood parameter estimates could be derived for each item type. The parameter estimates for recollection (R) and familiarity (as measured by d') are depicted in the upper panel of Table 5. The estimates for d' were 0.88 for practiced items (P+), 0.84 for unpracticed items (P-), and 0.83 for control items (C). The estimates for R were 0.55 for practiced items (P+), 0.17 for unpracticed items (P-), and 0.27 for control items (C).

Table 5: Parameter estimates for Experiment 3. Upper panel: Dual-process signal detection model. Lower panel: Unequal-variance signal detection model.

	Dual-process signal detection model				
	Parameter estimates		Goodness of fit		
Item Type	d'	R	χ^2	*df*	p
P+	0.88	0.55*	4.18	3	.24
C	0.83	0.27	2.17	3	.54
P-	0.84	0.17	1.46	3	.69
	Unequal-variance signal detection model				
	d'	σ	χ^2	*df*	p
P+	2.18*	1.57*	1.73	3	.63
C	1.30	1.26	0.32	3	.96
P-	1.12*	1.16	0.32	3	.96

Note: Asterisks indicate significant deviations from control performance ($p < .05$). P+ = practiced items; P- = unpracticed items; C = control items. d' denotes familiarity (dual-process model) or general memory strength (unequal-variance model), R denotes recollection, and σ denotes variance of the target distribution. *df* denotes the model's degrees of freedom.

Likelihood-ratio tests were conducted to examine whether the three types of items differed in their recollection and familiarity estimates. Regarding familiarity, the analysis showed no reliable change in parameter d' across conditions [$\chi^2(2) < 1$]. Regarding recollection, parameter R was significantly higher for practiced items (P+) compared to control items (C) [$\chi^2(1) = 17.24, p < .001$], but not significantly smaller for unpracticed items (P-) compared to control items (C) [$\chi^2(1) = 2.17, p > .10$].

Single-Process Analysis

The restrictive equal-variance signal detection model could describe the recognition ROCs for unpracticed items relatively well (P-) [$\chi^2(4) = 6.79, p > .10$], but had to be rejected for both the practiced and the control items (P+ and C) [$\chi^2(4)$'s> 29.00, p's$< .001$]. Therefore, the more general unequal-variance signal detection model was fitted to the data (see Appendix). The unequal-variance signal detection model explained the data very well for all three item types [all $\chi^2(3)$'s< 2.00, all p's$> .60$]. The goodness-of-fit statistics and the maximum-likelihood parameter estimates are depicted in the lower panel of Table 5.

Again, likelihood-ratio tests were used to examine whether the unequal-variance model's parameters varied across experimental conditions. The analysis showed that sensitivity as measured by d' was significantly higher for practiced items (P+) compared to control items (C) [$\chi^2(1) = 67.22, p < .001$], suggesting an increase in memory strength for practiced material. More interesting, parameter d' was reliably lower for unpracticed items (P-) compared to control items (C) [$\chi^2(1) = 8.46, p < .005$], suggesting a significant reduction in memory strength for unpracticed items. Furthermore, compared to control items (C), the variance of the old item's distribution as estimated by parameter σ was significantly larger for practiced material (P+) [$\chi^2(1) = 7.09, p < .01$], but not reliably smaller for unpracticed material (P-) [$\chi^2(1) < 1$]. The placement of the five confidence criteria did not significantly vary across conditions [$\chi^2(10) = 7.88, p > 0.60$].

Context Recall

On average, subjects correctly recalled the studied items' context colors for 70.7 % ($SE = 0.03$) of the practiced items (P+), for 65.7 % ($SE = 0.03$) of the control items (C), and for 67.2 % ($SE = 0.03$) of the unpracticed items (P-). Planned pairwise comparisons revealed significantly better context color recall for practiced (P+) compared to control items (C)[$t(63) = 2.09, p < 0.05, d = .22$]. However, no difference in context recall arose for unpracticed (P-) compared to control items (C) [$t(63) < 1$].

Discussion

Dual-process analysis of the data indicates a considerable increase in recollection for practiced items, but surprisingly no corresponding increment in familiarity. Further, the dual-process model indicates no reliable detrimental effects of retrieval practice on unpracticed material. If any, however, the model suggests a tendency for a decline in recollection, but no reduction in familiarity. Thus, dual-process analysis suggests that the retrieval practice effects in the present experiment using episodically related material are both qualitatively and quantitatively different from the effects obtained with semantically related material in Experiments 1 and 2. In particular, whereas in the previous experiments both the beneficial and the detrimental effects of retrieval practice were substantially familiarity-driven and statistically reliable, the effects in the present experiment were exclusively recollection-driven and the impairment of unpracticed material was not reliable. Unexpectedly however, regarding overall recognition performance, dual-process analysis indicates no generally enhanced role of recollective processes in the present item recognition task compared to the previous experiments.

The present findings further complicate the theoretical implications of a dual-process account of retrieval-induced forgetting: Whether retrieval-practice affects recollection and/or familiarity depends not only on the testing procedure (e.g., cued recall, associative recognition, or item recognition, see Experiments 1 and 2), but also on the structure of the study material (semantically or episodically related). The explanatory power of a dual-process account for the present results is additionally weakened when considering that subjects' ability to recall the perceptual context (i.e. color) of unpracticed items was not susceptible to retrieval-induced forgetting in the present experiment. In the dual-process framework, memory for the episodic context of studied episodes is assumed to be mainly recollection-driven (e.g. Mandler, 1980; Jacoby, 1991; Yonelinas, 2002). However, whereas dual-process analysis of the present item recognition data indicated at least a trend for a decline in recollection, not even a suggestion of such an effect was evident in context recall (C: 65.7%, P-: 67.2%).

Again, the unequal-variance single-process model led to a technically better description of the recognition data than the dual-process model. In contrast to the dual-process model's familiarity measure, the single-process estimates for general memory strength exhibited not only a reliable increase for practiced material, but also a significant decrement for unpracticed material, and thus align with the pattern of results obtained in Experiments 1 and 2. Thus, whereas dual-process theory suggests fundamentally different consequences of retrieval practice depending on the structure (semantic categories versus perceptual categories) of the studied material, the single process analyses of Experiments 1-3 converge on the view that retrieval practice leads to a comparable modulation in general memory strength in both experimental contexts.

The present finding of no decrease in subjects' ability to remember the unpracticed items' perceptual context ostensibly conflicts with Ciranni& Shimamura's (1999) study which suggests that the forgetting generalizes even to features of the unpracticed material that were not retrieval practiced. However, in Ciranni and Shimamura (1999) such features (e.g. shape) were attributes of the presented items 'themselves', whereas in the present experiment the colors were attributes of the perceptual context against which the items (i.e. words) were presented. Research on feature binding in episodic memory suggests that an object's color is easily *bound* into a single memory representation of a colored object, whereas presenting items against colored backgrounds rather creates associative links between the objects' representations and the respective background colors (e.g. Ecker at al., 2004; Groh-Bordin et al., 2006). Assuming that retrieval practice directly weakens the unpracticed items' memory representations, only features that are inherent parts of these representations might be susceptible to retrieval-induced forgetting. The empirical validity of such a proposal will be examined in Experiment 4.

Experiment 4

To explore the role of episodic feature binding in retrieval-induced forgetting, a variation of the experimental design used in Experiment 3 was employed. In Experiment 4, subjects studied semantically unrelated items displayed in different font colors. Thus, the episodic grouping factor *color* was presented as a perceptual feature of the study items 'themselves', whereas in the previous experiment, the colors were features of the items' perceptual context. Again, during retrieval practice, a subset of the items from half of the artificially created color categories was repeatedly retrieved. Finally, a color recognition test was applied where all items from the study list were presented either in their original font color or in a different color. Again, the ROC-procedure was used to investigate the qualitative effects of retrieval practice on recognition memory. Assuming that the presentation of words in different font colors creates 'bound' episodic representations of colored items in memory, the retrieval practice effect should not only modulate item recognition performance (see Experiment 3), but also recognition of the color features bound to these memory representations.

Method

Participants

Subjects were 72 young adults (19-34 years old) who were paid six Euro for participating in the experiment. They were tested individually in sessions that lasted approximately 25 minutes.

Material

The primary stimuli were 32 semantically unrelated German words chosen from the norms reported in Mannhaupt (1983). None of the items was pre-experimentally associated with any particular color.

Design

The experiment consisted of three main phases: A study phase, a retrieval practice phase and a color recognition test phase. The words were arbitrarily divided into four sets of eight items. For each subject, the font colors red, green, blue, and yellow were randomly assigned to the four "categories" in the study phase. From each of two of the "color categories", four target items were used in the retrieval practice phase. Accordingly, eight items (displayed in two colors, e.g. red and blue) were retrieval practiced (P+), eight items (displayed in the same colors as P+) were unpracticed (P-), and 16 Items (displayed in the two remaining colors, e.g. green and yellow) served as control items (C). Assignment of the items to retrieval practice was counterbalanced, so that every item was practiced equally often across subjects. All studied items were used in the recognition test phase, where half of the items was displayed in the same font color as they were studied (i.e. target, correct color), whereas the other half was displayed in any other of the four font colors (i.e. lure, incorrect color). The assignment of the items to correct vs incorrect font color at test was counterbalanced and it was assured that for each item type (P+, P-, C) equally many items served as targets and lures.

Procedure

Subjects were tested individually in a quiet surrounding, comfortably seated in front of a 15" TFT Computer screen. They were instructed to memorize each item's color on the study list and that memory for the items' colors only would be tested later on, as all words would be shown again at test in eventually different colors. Every item was presented for 1500 ms in its assigned font color, followed by a 500 ms blank screen. The serial order of the items was blocked randomized, with the restriction that no item was displayed in the same color as its antecessor. Furthermore, three buffer items were added at the beginning and at the end of the list and the buffer items were also randomly assigned to one of the four font colors. Following the study phase, subjects had to count backwards from 300 in steps of threes for approximately 60 seconds.

Subsequently, the retrieval practice phase started. The word stems were displayed in the same color as the corresponding items in the study phase. Subjects were instructed to use the font colors as cues to complete the word-stems from memory. The serial order of the word-stems was quasi-random, with alternating font colors. The experimenter noted the subjects response on a prepared data sheet and the subjects could proceed to the next item by pressing a key. After retrieval practice, subjects had to solve a set of simple arithmetic problems for about five minutes and were allowed to use a pocket calculator to avoid frustration in the subject.

In the final color recognition test phase, each test item was presented in the middle of the screen, together with a schematically depicted 6-point rating scale ranging from *1: definitely old color* to *6: definitely new color* in the lower part of the screen. Subjects were instructed to enter their responses via the digits of the PC-Keyboard. As soon as any numerical response was entered, the next item was presented on the screen; subject's responses were automatically recorded in a log file. The serial order of the test items was randomized in the same manner as in the previous experiments.

Results

Retrieval-Practice Phase

Participants, on average, successfully completed 59.5% (SE = 0.02) of the wordstems presented during retrieval practice.

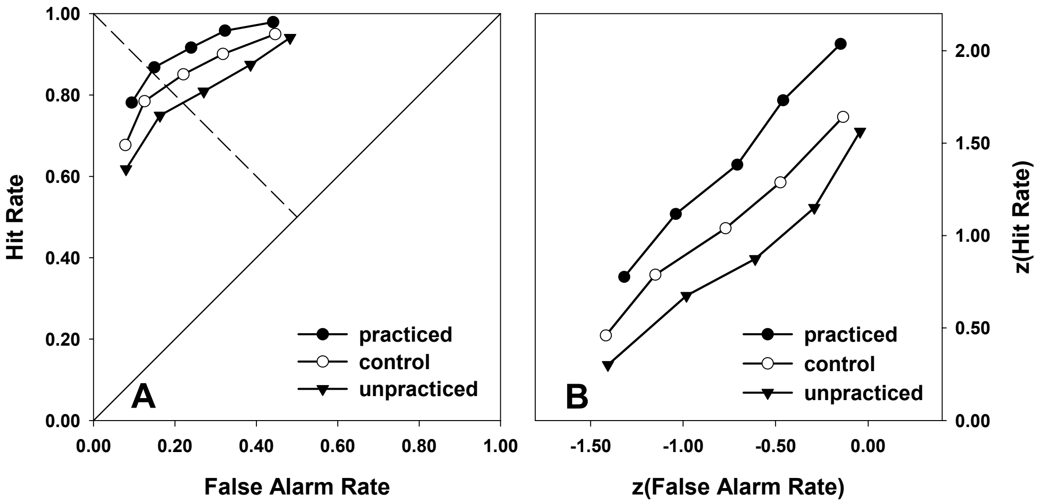

Figure 3: Recognition ROCs (A) and zROCs (B) for the three item types in Experiment 4.

Recognition Test

The color discrimination ROCs obtained by plotting the cumulative false alarm rates against hit rates for each of the three item types are illustrated in Figure 3.

Dual-Process Analysis

The dual-process model (Yonelinas, 1994) described the data set well [all $\chi^2(3)$'s$<$ 1.70, all p's$>$ .60]. The parameter estimates for familiarity (as measured by d') and recollection (R) are given in the upper panel of Table 6. Although by tendency, the values for d' (i.e., familiarity) exhibited the typical pattern of retrieval-induced forgetting (P+ $>$ C $>$ P-), likelihood-ratio tests revealed neither a significant increase of that measure for practiced items (P+) compared to control items (C) [$\chi^2(1) = 2.84, p > .05$] nor a significant decrease for unpracticed items (P-) compared to control items (C) [$\chi^2(1) = 1.31, p > .20$]. Parameter R (recollection) varied unsystematically across item types and the changes in the parameter estimates were not statistically reliable [P+ vs C: $\chi^2(1) < 1$, P- vs C: $\chi^2(1) < 1$][4]. Across conditions, the estimates for recollection (R) were not significantly different from zero [all $\chi^2(1)$'s< 1.80, all p's$> .15$], indicating that recognition was predominantly - if not exclusively - based on familiarity.

Single-Process Analysis

The unequal-variance signal detection model described the recognition data very well [all $\chi^2(3) < 2.10$, all $p > .55$] (see middle panel of Table 6). Although by tendency, parameter d' exhibited the standard pattern of retrieval-induced forgetting (P+ > C > P-), likelihood-ratio tests revealed no significant variations of this measure across item types [$\chi^2(2) = 2.15, p > .30$]. Across conditions, the estimates for parameter σ were relatively close to 1, indicating equal variances of the underlying strength distributions.

Therefore, the more restrictive equal-variance signal detection model was fitted to the data (assuming $\sigma = 1$, see lower panel of Table 6). The model described the data well [all $\chi^2(4) < 3.40$, all $p > .50$] and its goodness of fit was not statistically different from that of the unequal-variance signal detection model [$\chi^2(3) < 3.16, p > .35$]. The equal-variance model's estimates for d' exhibited the standard pattern of retrieval-induced forgetting (see lower panel of Table 6): P+ (2.14) > C (1.84) > P- (1.57). Statistical analysis shows that general memory strength as measured by d' was significantly higher for practiced (P+) items [$\chi^2(1) = 14.43, p < .001$], and reliably lower for unpracticed (P-) items compared to control (C) [$\chi^2(1) = 15.94, p < .001$].

[4]The parameter estimates for P+ condition might be distorted due to unusually high performance levels (see Yonelinas, 2002) and thus should not be interpreted in detail. However, performance levels for P- and C conditions were in an acceptable range and thus most likely yielded meaningful parameter estimates

Table 6: Parameter estimates for Experiment 4. Upper panel: Dual-process signal detection model. Middle panel: Unequal-variance signal detection model. Lower panel: Equal-variance signal detection model.

	Dual-process signal detection model				
	Parameter estimates		Goodness of fit		
Item Type	d'	R	χ^2	*df*	p
P+	2.14	0.00	0.44	3	.93
C	1.62	0.27	1.04	3	.79
P-	1.30	0.29	1.68	3	.64
	Unequal-variance signal detection model				
	d'	σ	χ^2	*df*	p
P+	2.05	0.93	0.26	3	.97
C	1.98	1.14	0.91	3	.82
P-	1.72	1.17	2.03	3	.57
	Equal-variance signal detection model				
	d'	σ	χ^2	*df*	p
P+	2.14*	1 (fixed)	0.44	4	.98
C	1.84	1 (fixed)	2.58	4	.63
P-	1.57*	1 (fixed)	3.34	4	.50

Note: Asterisks indicate significant deviations from control performance ($p < .05$). P+ = practiced items; P- = unpracticed items; C = control items. d' denotes familiarity (dual-process model) or general memory strength ((un)equal-variance model), R denotes recollection, and σ denotes variance of the target distribution. *df* denotes the model's degrees of freedom.

Discussion

The present experiment demonstrates retrieval-induced forgetting in an experimental setting that differed from the standard retrieval-practice paradigm in several aspects. Like in Experiment 3, verbal stimuli were grouped by an episodic feature (i.e., color) instead of using semantic categories. In Experiment 3, using context color as grouping factor, retrieval practice impaired item recognition of unpracticed material but left memory for these items' episodic contexts unaffected. In the present experiment, however, using font color as grouping factor, retrieval practice on a subset of the words reliably impaired episodic recognition of the unpracticed words' colors. On the basis of these findings, it is argued that the studied items' font colors were 'bound' into unitary memory representations of colored items.

In contrast to the foregoing experiments, using tests of item recognition, an episodic feature discrimination task was employed in the present experiment. In some respect, memory for an item's perceptual features (i.e., font color) can be considered as associative information. Accordingly, assuming a dual-process framework of recognition memory (e.g. Yonelinas, 2002), the assessment of such information at test should be largely based on episodic recollection. Interestingly, dual-process analysis of the ROC data from the present experiment yielded a fundamentally different picture: Color discrimination was largely based on familiarity and hardly involved recollection processes at all.

However, these unexpected results need not necessarily invalidate the dual-process model's conceptualization of recollection as an episodic, associative process. It appears well possible that the particular recognition task employed in the present experiment may have rendered the use of recollection processes obsolete in the present experimental situation. Because font color may have been memorized as an integral part of the studied items, rather than as associative or context information, episodic recognition of font color could indeed have been based on stimulus familiarity rather than on recollection (see Eichenbaum & Bunsey, 1995; Quamme, 2004 or Yonelinas et al., 1999 for similar considerations). Consistent with such a proposal, Groh-Bordin et al. (2006) showed that changing the colors of studied objects in tests of item recognition affected the FN400 ERP-component which is putatively related to familiarity, whereas changing the colors of the objects' perceptual contexts modulated the late parietal component which is putatively related to recollection.

Fitting the dual-process model to the color recognition data yielded no significant decreases in recollection or familiarity for the unpracticed material. By tendency however, the forgetting in the present experiment was clearly familiarity-driven, consistent with the finding of reduced familiarity for unpracticed material in Experiments 1 and 2. Indeed, because color recognition appeared to rely exclusively on familiarity, it can not be determined from the present analysis to what extent recollection might have been affected as well. Still, from a dual-process perspective, the results may suggest that retrieval-induced forgetting of episodically related material can affect not only recollection (Experiment 3), but also familiarity.

Assuming that recollection did not significantly contribute to recognition performance, the dual-process model technically reduces to an equal-variance single-process model, rendering the opposing models' accounts of the present data empirically indistinguishable. From a single-process perspective, the present results again indicate that retrieval-induced forgetting affects the unpracticed material's general memory strength, consistent with the results from Experiments 1-3. Moreover, consistent with Ciranni & Shimamura's (1999) finding of generalized forgetting in episodic recall, such reductions in memory strength can generalize to recognition of episodic features that are 'bound' into the unpracticed items' memory representation.

Interim Summary

In Experiments 1-4, the effects of retrieval practice on recognition performance were examined on the basis of dual- and single-process models of recognition memory. The dual-process results for Experiments 1-4 are summarized in Figures 4A and 4B. In Experiments 1 and 2, using semantically related words in tests of item recognition, the forgetting of unpracticed material was predominantly familiarity-driven, indicating that familiarity indeed plays an important role in retrieval-induced forgetting. In Experiments 3 and 4, using perceptually related material, a mixed picture arose: In item recognition (Experiment 3), the forgetting was predominantly recollection-driven, whereas in feature recognition (Experiment 4), it again was primarily familiarity-driven. Thus, whether or not retrieval-induced forgetting affects recollection and/or familiarity seems to depend in a complex manner on both the structure of the study material (semantic versus episodic categories) and the type of test (item versus feature recognition). However, contrary to expectations based on prior work, the reductions in recollection appeared to be generally less robust than the pronounced decreases in familiarity.

The single-process results for Experiments 1-4 are summarized in Figure 4C. In contrast to the complex outcomes implied by the dual-process models, single-process analysis of the present data yielded an invariant pattern of significant modulations in parameter d' across experiments, regardless of various differences in material and procedure. From a single-process perspective, retrieval practice essentially reduced the unpracticed material's general memory strength, consistent with the view that retrieval inhibition weakens the unpracticed material's memory representation. In all, standard and generalized signal detection models led to an excellent description of the recognition data and appeared to be conceptually superior to the dual-process models (see Experiments 1-3). In particular, the single-process interpretation provides a more parsimonious account of the present results than a dual-process interpretation.

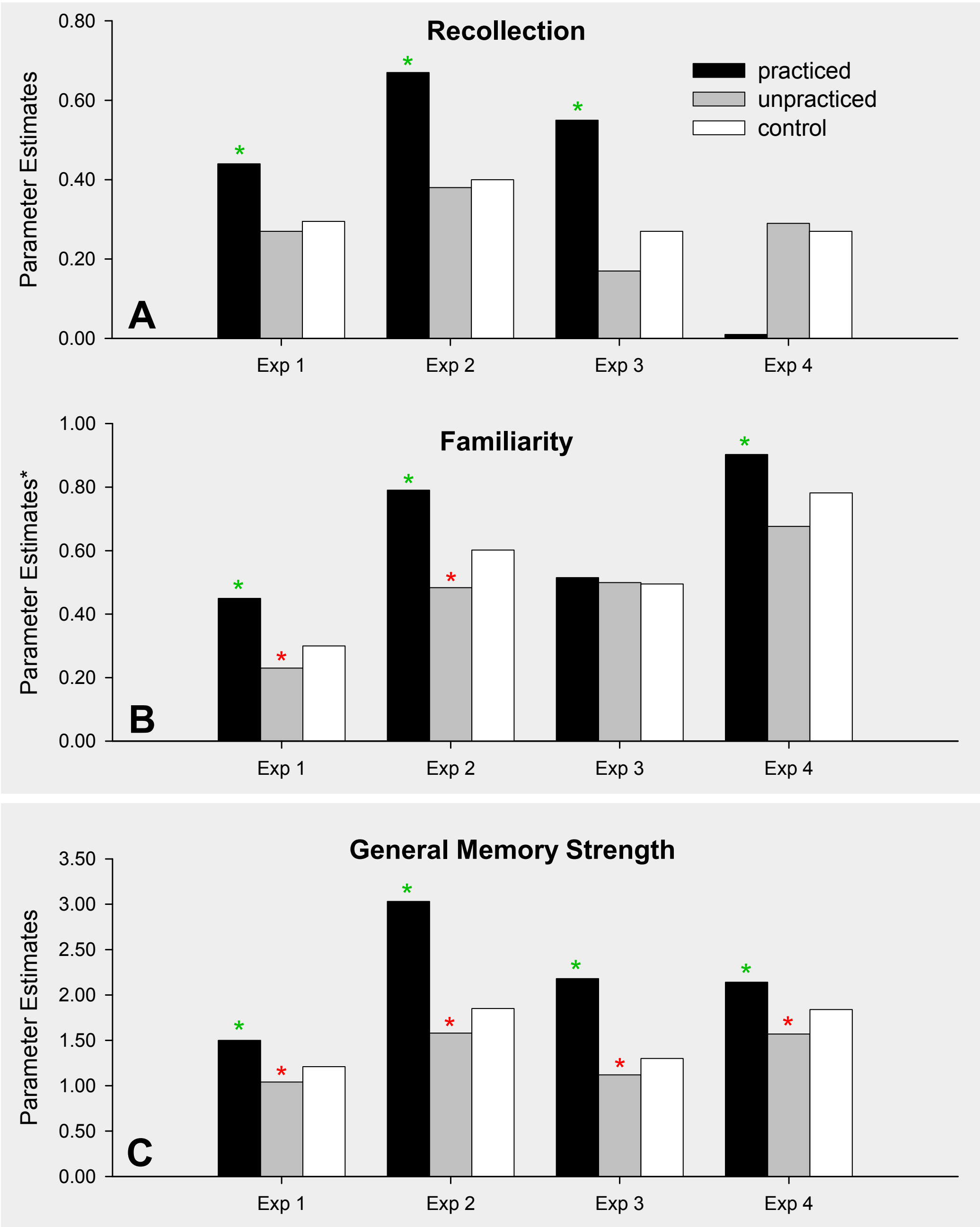

Figure 4: Summary of the results from Experiments 1-4. A: Dual-process estimates for recollection. B: Dual-process estimates for familiarity. *For Experiments 2-4, d''s were converted into probabilities for familiarity (see Yonelinas, 2002). C: Single-process estimates for general memory strength derived from the unequal-variance signal detection model (Experiments 1-3) or from the equal-variance signal detection model (Experiment 4). Green and red asterisks indicate significant deviations from control performance (green: increase, red: decrease, $p < .05$).

Part III

Electrophysiological Correlates of Retrieval-Induced Forgetting and Recognition Memory

Experiment 5

The main goal of Experiment 5 was to investigate the electrophysiological correlates of retrieval-induced forgetting in item recognition. The analysis of EEG old/new effects (see Background section) should further elucidate whether the detrimental effects of retrieval practice on unpracticed memories are mainly due to decreases in recollection as suggested on the basis of prior work (Anderson, 2003; Verde 2004), or whether the forgetting affects the familiarity of the unpracticed material as well. Indeed, Experiments 1, 2, and 4, using behavioral models to assess recollection and familiarity, yielded strong evidence for retrieval-induced decreases in familiarity. However, behavioral dual-process models are highly disputed in the literature (Dunn, 2004; Rotello et al., 2004; Wixted, 2007; Wixted & Stretch, 2004) and yielded a theoretically unbalanced description of the present work's behavioral data. Thus, these models might have produced a distorted picture of the specific effects of retrieval-induced forgetting in recognition memory.

Except for modifications required for EEG recording during item recognition, Experiment 5 was methodologically highly similar to Experiments 1 and 2, using semantically related study material in a standard retrieval-practice paradigm. Like in Experiment 1, the remember/know procedure was used, thus allowing for simultaneous evaluation of behavioral and electrophysiological measures of recollection and familiarity. On the basis of the previous results, it may be expected that retrieval practice affects both behavioral estimates and EEG correlates of the unpracticed material's familiarity. In contrast, for practiced material, an increase in both processes should arise across both measurement methods.

It is widely assumed that familiarity is characterized by early/frontal ERP old/new effects (FN400 old/new effects), whereas late/parietal old/new effects reflect episodic recollection (for reviews, see Curran et al., 2006; Friedman & Johnson, 2000; Rugg & Yonelinas, 2003; Woodruff et al., 2006; but see also McKenzie & Donaldson, 2007, for differing topographies). However, this functional dissociation is not without counterarguments (Finnigan et al., 2002; Voss & Paller, 2006; Yovel & Paller, 2004). Using the remember/know procedure, the present experiment allows for a two-fold investigation of the dual-process interpretation of the present findings: On the one hand, the expected pattern of behavioral dual-process results can be compared to the effects on the putative EEG correlates of recollection and familiarity across item type, in order to test the internal consistency of the neuro-cognitive dual-process framework. Further, the EEG correlates of remembering and knowing will be analyzed separately to determine whether the assumed functional dissociation of old/new effects across response categories replicates in the present experimental context.

Until recently, recognition memory research has mainly focused on analysis of old/new differences in the waveforms of event-related potentials (ERP old/new effects, see Background section). However, a number of studies using measures of event-related (de)synchronsiation (ERD/ERS, see Background/Method section) have further identified oscillatory recognition correlates in the theta frequency range (4-7 Hz, e.g., Klimesch et al., 2000; 2001; 2006; Jacobs et al., 2006). In the present experiment, both ERP and ERD/ERS, but also measures of inter-electrode synchronisation (phase locking value (PLV), see Method section) were employed to investigate possible effects of retrieval-induced forgetting and to explore the electrophysiological correlates of recognition memory in more detail. Inter-electrode synchronisation of brain oscillations is assumed to reflect enhanced interaction between distant brain regions in terms of increased information exchange (e.g., Fries, 2005; Lachaux et al., 1999; Varela et al., 2001). In prior work, such inter-electrode synchronisation of oscillations in the theta frequency range (4-7 Hz) has been associated with binding mechanisms in episodic memory (Summerfield & Mangels, 2005) and might thus also play an important role in episodic recognition.

Method

Participants

Subjects were 30 adults (19-34 years old, 16 female, 14 male) who participated voluntarily in the experiment. They were tested individually in two subsequent sessions that each lasted approximately 30 minutes. Five subjects were excluded from the data analysis because they produced less than 15 artifact-free trials in at least one of the experimental conditions or response categories.

Material

For each session, stimuli were 12 concrete German words from each of nine semantic categories (Battig & Montague, 1969; Mannhaupt, 1983). In session one, the categories were *body part, sport, musical instrument, quadruped, piece of furniture, tool kit, spice, tree,* and *car part.* In session two, the categories were *alcoholic drink, fruit, organ, bird, disease, article of clothing, african state, planet,* and *profession.* The chosen exemplars were 12 relatively strong exemplars of the respective category (rank 4-15 in the norms). Within each of the nine categories, six of the chosen examplars (rank 7-12) were studied, and the remaining six (rank 4-6 and 13-15) were used as lures in the recognition test. Additionally, four exemplars from each of the three categories *hygiene, sweets,* and *stationery* were used as buffer items in session one, and four exemplars from each of the three categories *media, pasta* and *geometric shapes* were used as buffer items in session two, for a total of 12 buffer items in each session.

Design

The experiment consisted of two identical sessions. For each of the two sessions, the experimental design was identical to Experiment 2, with the following exceptions: In the study phase, 54 exemplars from nine categories were studied. In the retrieval practice phase, subjects practiced three exemplars from each of six of the nine studied categories. Thus, in each session, 18 items were practiced (P+), 18 items from practiced categories were unpracticed (P-), and 18 items from the remaining three categories served as controls (C). In the final test phase, a sequential remember/know testing procedure was conducted.

Procedure

Subjects were tested individually in a quiet surrounding, seated in front of a 15" computer screen. At the beginning of the study phase, an instruction to memorize all to-be-presented words was displayed. Then, each item was presented for 2000 ms, followed by a 500 ms blank screen. Throughout the experiment, all items were displayed in the middle of the screen in a white Arial font (boldface pts 20) against a black background. The order of the items was randomized in the same manner as in the previous experiments. Additionally, four buffer items were shown at the beginning of the study list, and four were shown at the end. After the study phase, subjects were instructed to count backwards from 500 in steps of threes for approximately 60 seconds.

Subsequently, the retrieval-practice phase started in which subjects practiced 18 (6×3) of the studied items. A category/word-stem pair was presented on the screen (e.g. FRUIT - *Ap...*) and subjects were instructed to complete the word-stem with a studied item. The experimenter noted the subject's response on a prepared data sheet and participants could proceed to the next item by pressing a key. The order of the category-wordstem pairs was randomized in the same manner as in Experiments 1 and 2. For the following five minutes, a set of simple arithmetic problems was presented. Participants were allowed to use a pocket calculator.

In the final test phase, a sequential remember/know-recognition test was conducted. Subjects 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. Each trial consisted of a 350 ms blank screen, followed by a 2200 ms fixation cross, then the test item was presented for 1200 ms, followed by a 500 ms blank screen. Subsequently, the question ”new (n) or old (o)?” was displayed and subjects were instructed to enter their response via specified keys on the PC keyboard. In case of a ”new” response, the next test trial started immediately. In case of an ”old”-response the question ”know(k) or remember(r) ?” was presented on the screen, and subjects had to enter a response by pressing the corresponding key on the PC Keyboard before proceeding to the next trial. The order of the items was block randomized in the same manner as in the previous experiments. After completion of the first session and following a break of about ten minutes, the procedure was identically repeated with the second set of item material.

EEG Recording

During the recognition test phase, EEG activity was recorded continuously from 63 Ag/AgCl electrodes mounted on a elastic cap according to the extended 10- 20 system. Signals between 0.3 and 30 Hz were recorded and digitized at a sample rate of 500 Hz using Brain Amp MR (Brain Products, Munich; all impedances kept below 5 kΩ). Recordings were initially referenced to FCz, then converted to an average reference off-line. BESA 5.1.6 (Brain Electrical Source Analysis MEGIS Software) was used to remove eye movements, blinks, and other artifacts from the continuous EEG data. The recording was then segmented in single trials having a length of 4000 ms (starting 2000 ms before the onset of each test item). Trials containing muscle, movement, or other artifacts were manually rejected.

ERP Analysis

Inferred from recognition literature reporting early (<500 ms) ERP old/new effects at frontal recording sites and late (>500 ms) ERP old/new effects at parietal recording sites (e.g. Rugg & Yonelinas, 2003), recognition ERPs were analyzed for two regions of interest (ROIs; Frontal: Fz, F1, F2, F3, F4, FCz, FC1, FC2, FC3, FC4; parietal: Pz, P1, P2, P3, P4, CPz, CP1, CP2, CP3, CP4) and for two time windows (early: 250-500 ms; late: 500-750 ms) that were based on previous studies, while taking into account visual inspection of the waveforms.

ERD/ERS Analysis

The ERD/ERS is defined as percentage power decrease (ERD) or power increase (ERS) in relation to a reference interval. In the present study, a reference interval of 500 ms was chosen (-750 ms to -250 ms before stimulus onset). By means of temporal spectral evolution (TSE) analysis (Hari & Salmelin, 1997), ERD/ERS in the theta and upper alpha frequency domains (4-7 Hz and 10-13 Hz) were calculated as implemented in BESA 5.1.6 (Brain Electrical Source Analysis, MEGIS Software). Based on previous evidence (Klimesch et al., 2001; 2006) while also taking into account visual inspection of the data, the theta ERS data were statistically analyzed for slightly different time windows than the ERP data (early: 200-400 ms, late: 500-700 ms). The upper alpha ERD data were analyzed for a later time window (750-1100 ms) only.

Phase Locking Value (PLV) Analysis

For PLV-analysis, trial numbers for each subject were matched across conditions and response categories by random selection. The PLV estimates the phase coupling of oscillations in a certain frequency range between two electrode sites. Using the software BESA 5.1.6 (Brain Electrical Source Analysis

MEGIS Software), the PLV was calculated for each electrode pair for a frequency range of 4-20 Hz with a frequency resolution of 1 Hz. The PLV was collapsed over three time intervals of interest (200-500, 500 -800, and 800-1200 ms after stimulus onset) and analyzed separately for three frequency bands of interest (theta: 4-7 Hz, lower alpha: 7-10 Hz, and upper alpha: 10-13 Hz). Prior to statistical analysis, the PLV was Fisher-z-transformed. For statistical analysis of the PLV, a two-stage procedure was carried out. At first, t-tests were calculated for each electrode pair to investigate which electrode pairs showed a significant difference between the two conditions ($p <$ 0.005; one-tailed). Second, a randomization test (Blair & Karniski, 1993), based on 20 000 permutation runs, was carried out to control for type I errors due to multiple testing. This procedure evaluates whether a given number of electrode pairs, exhibiting a significant difference between the two conditions (remember vs. forget), is expected by chance. If the p-value of this randomization test falls below .05, less than 5% of the permutation runs exhibited equal or more electrode pairs with a significant difference between the two conditions. Only results for frequency bands and time windows which show a p-level of smaller than .05 in the randomization test will be reported.

Behavioral Results

Retrieval-Practice Phase

In the retrieval-practice phase, participants, on average, successfully completed 89.4% ($SE = 0.01$) of the category/word-stem pairs.

Recognition Performance

The recognition data are depicted in Table 7. Recognition and remember hit and false alarm rates were examined separately for practiced (P+) and unpracticed (P-) items.

Table 7: Hit rates and false alarm rates in Experiment 5. Upper panel: Old/new recognition performance. Lower Panel: Remember performance.

	Hits		False Alarms	
Item Type	Rate	*SE*	Rate	*SE*
		Recognition		
P+	0.96*	0.01	0.19	0.03
C	0.85	0.01	0.21	0.03
P-	0.77*	0.02	0.19	0.03
		Remember		
P+	0.71*	0.03	0.04	0.01
C	0.53	0.03	0.03	0.01
P-	0.48*	0.02	0.04	0.01

Note: Asterisks indicate significant deviations from control performance ($p < .01$). P+ = practiced items; C = control items; P- = unpracticed items; SE = standard error.

The recognition hit rate was .96 for practiced (P+) items and .85 for control items (C), the remember hit rate was .71 for practiced items and .53 for control items. A 2 $\times$ 2 analysis of variance with the two factors of item type (practiced vs. control) and criterion (old/new vs. remember) showed a significant main effect of item type [$F(1, 24) = 50.18$, $MSE = 0.010$, $p < .001$, partial $\eta^2 = .68$], and a significant main effect of criterion [$F(1, 24) = 161.16$, $MSE = 0.012$, $p < .001$, partial $\eta^2 = .87$]. These effects reflect more hits for practiced items (P+) than for their controls (C+) and more frequent recognition than remember hits. The interaction between retrieval status and response criterion was also significant [$F(1, 24) = 5.39$, $MSE = 0.005$, $p < .05$], reflecting a stronger increase in remember than in recognition hits for practiced compared to control items.

The recognition hit rate was .77 for unpracticed items (P-) and .85 for control items (C-), the remember hit rate was .48 for unpracticed items and .53 for their controls. A 2 $\times$ 2 analysis of variance with the two factors of item type (unpracticed vs. control) and criterion (old/new vs. remember) yielded a significant main effect of item type [$F(1, 24) = 13.68$, $MSE = 0.009$, $p < .01$, partial $\eta^2 = .36$], and a significant main effect of criterion [$F(1, 24) = 290.04$, $MSE = 0.008$, $p < .001$, partial $\eta^2 = .92$], reflecting fewer hits for unpracticed items (P-) than for their controls (C-) and again more frequent recognition than remember hits. There was no reliable interaction between the two factors [$F(1, 24) = 1.21$, $MSE = 0.003$, $p > .25$].

False alarm rates were examined separately. As expected, a 2 $\times$ 2 analysis of variance with the two factors of category type (practiced vs. control) and criterion (old/new vs. remember) showed a significant main effect of criterion [$F(1, 24) = 48.11$, $MSE = 0.014$, $p < .001$, partial $\eta^2 = .67$], reflecting the fact that recognition false alarms are far more frequent than remember false alarms. Besides, no significant main effect of category type [$F(1, 24) = 1.40$, $MSE = 0.002$, $p > .20$] was found. The interaction between the two factors was significant [$F(1, 24) = 4.66$, $MSE = 0.001$, $p < .001$, partial $\eta^2 = .16$], reflecting that the recognition false alarm rate for exemplars from practiced categories (P) was marginally smaller than for exemplars from unpracticed categories (C) [$t(24) = 1.81, p < 0.1$], while the remember false alarm rates were almost equal for both item types [$t(24) < 1$].

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 (for technical details, see Experiment 1). The process estimates are depicted in the upper panel of Table 8.

The recollection estimates were .67 for practiced items (P+) and .50 for controls (C). The familiarity estimates were .69 for P+ items and .48 for C items. A 2 × 2 analysis of variance with the two factors of item type (practiced vs. control) and process estimate (recollection vs. familiarity) showed a significant main effect of item type [$F(1, 24) = 85.38, MSE = 0.011, p <$ .001, partial η^2 = .78], indicating a general memory improvement for the practiced material. The estimates for recollection and familiarity did not differ significantly [$F(1, 24) < 1$], and there was no interaction between the two factors [$F(1, 24) = 1.28, MSE = 0.009, p > .25$]. Planned comparisons revealed a significant increase in both recollection [$t(24) = 5.33$, $p < .001$, two-tailed] and familiarity [$t(24) = 8.65$, $p < .001$, two-tailed] of the practiced items.

The recollection estimate for unpracticed items (P-) was .44 compared to .50 for controls (C). The familiarity estimates were .41 for P- items and .48 for C items. A 2 × 2 analysis of variance with the two factors of item type (unpracticed vs. control) and process estimate (recollection vs. familiarity) showed a significant main effect of item type [$F(1, 24) = 14.63, MSE = 0.007, p < .01$, partial $\eta^2 = .38$], indicating generally impaired memory processes for the unpracticed material. The estimates for recollection and familiarity did not differ significantly [$F(1, 24) < 1$], and there was no interaction between the two factors [$F(1, 24) < 1$]. Planned comparisons revealed a significant decrease in both recollection [$t(24) = 2.73$, $p < .05$, two-tailed] and familiarity [$t(24) = 2.56$, $p < .05$, two-tailed] of the unpracticed items.

I also fitted the STREAK model (Rotello, Macmillan, & Reeder, 2004) to the remember/know-data (see Experiment 1 for technical details). The best fitting STREAK parameters for both global (d_x) and specific (d_y) memory strength are shown in the middle panel of Table 8. For practiced (P+) items, the increase in global as well as in specific memory strength was statistically reliable [d_x: $\chi^2(1)$= 43.45, $p < .001$; d_y: $\chi^2(1)$= 31.11, $p < .001$]. For unpracticed items (P-), both global and specific memory strength were significantly reduced [d_x: $\chi^2(1)$= 4.28, $p < .05$; d_y: $\chi^2(1)$= 6.07, $p < .05$].

Single-Process Analysis

Like in Experiment 1, the equal-variance signal detection model was fitted to the remember/know data. The model described the data well for three of the four item types [all $\chi^2(1)'s < 1.90$, $p > .15$], but had to be rejected for practiced (P+) items [$\chi^2(1) = 5.97$, $p < .05$]. Therefore, the unequal-variance signal detection model was fitted to the data. When directly compared with the more restrictive equal-variance model, the unequal-variance model described the data set significantly better [$\chi^2(3)$= 8.90, $p < .05$], 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 8).

The parameter estimates showed a marginally higher d' for practiced items (P+) compared to their controls (C+) (2.22 vs. 1.96) [$\chi^2(1)$= 3.40, $p < .10$], suggesting an increase in memory strength of the practiced material. They further showed marginally lower d' for unpracticed items (P-) compared to their controls (C-) (1.73 vs. 1.96) [$\chi^2(1)$= 3.33, $p < .10$], indicating a reduction in the unpracticed material's memory strength. Parameter σ was significantly smaller for practiced compared to control items [$\chi^2(1)$= 5.81, $p < .05$], but no difference in σ emerged for unpracticed compared to control items [$\chi^2(1) < 1$]. The placement of remember and know criteria

Table 8: Parameter estimates for Experiment 5. Upper panel: Independence remember/know model. Middle Panel: STREAK model. Lower Panel: Unequal-variance signal detection model.

	Independence Remember/Know			
Item Type	F	R		
P+	.69*	.67*		
C	.48	.50		
P-	.41*	.44*		
	STREAK			
Item Type	d_x	d_y	C_0	C_r
P+	1.41*	2.38*	0.72	0.63
C	0.94	1.80	0.65	0.33
P-	0.81*	1.56*	0.72	0.31
	Unequal-Variance Signal Detection Model			
Item Type	d'	σ	k	r
P+	2.22$^\diamond$	0.78*	0.89	1.81
C	1.96	1.13	0.81	1.87
P-	1.73$^\diamond$	1.15	0.89	1.81

Note: Asterisks indicate significant deviations from control performance ($*: p < .05$, $\diamond : p < .10$). P+ = practiced items; C = control items; P- = unpracticed items. F denotes familiarity, R recollection. d_x denotes global memory strength, d_y specific memory strength, C_0 location of the STREAK old/new criterion, C_r location of the STREAK remember criterion, d' general memory strength, σ variance of the target distribution, k location of the know criterion, and r location of the remember criterion.

(parameters r and k) did not vary with item type [$\chi^2(4)$= 3.17, $p > .50$].

Electrophysiological Results A: Retrieval-Induced Forgetting

For analysis of the electrophysiological correlates of retrieval-induced forgetting, hit trials for each item type were collapsed across remember- and know response categories. For investigation of memory-related EEG effects, overall hit trials for each studied item type (P+, C, and P-) were contrasted with correct rejections of unstudied items (new).

ERP Results

The grand average ERPs pooled over parietal recording sites for each item type are shown in Figure 5. A 4 × 2 × 2 analysis of variance with the three factors of item type (P+, C, P-, New), ROI (frontal vs parietal), and time window (early vs late) revealed a significant triple interaction [$F(3, 72) = 6.04, MSE = 0.127, p < .001$]. Subordinate analyses revealed no effect of item type and thus no reliable old/new effects at frontal recording sites in the early time window (FN400) [$F(3, 72) < 1$], but significant old/new effects emerged at parietal recording sites in the late time window [$F(3, 72) = 11.99, MSE = 0.236, p < .001$]. Planned comparisons revealed a significantly stronger late/parietal old/new effect for practiced (P+) compared to control items (C) [$t(24) = 2.93, p < .005$]. However, the trend for a smaller old/new effect for unpracticed (P-) compared to control items (C) was not reliable [$t(24) < 1$].

Theta ERS Results

The pooled theta ERS data for each item type are shown in Figure 6A. A 4 × 2 × 2 analysis of variance with the three factors of item type (C, P+,

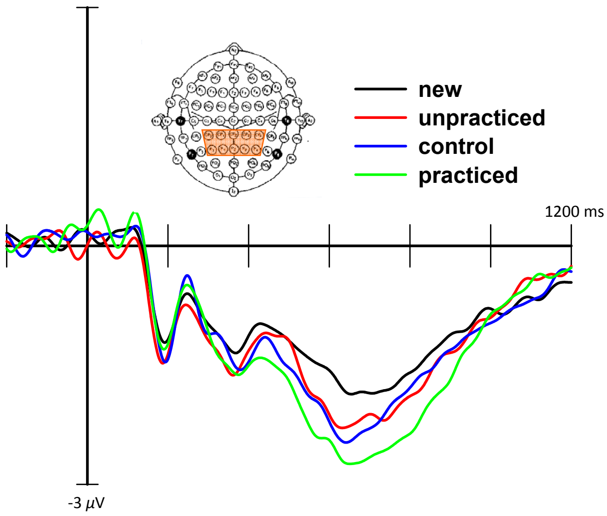

Figure 5: Grand average ERPs pooled over parietal recording sites for each item type, Experiment 5. New: Correct rejections; Unpracticed: Hits to unpracticed items; Control: Hits to control items; Practiced: Hits to practiced items.

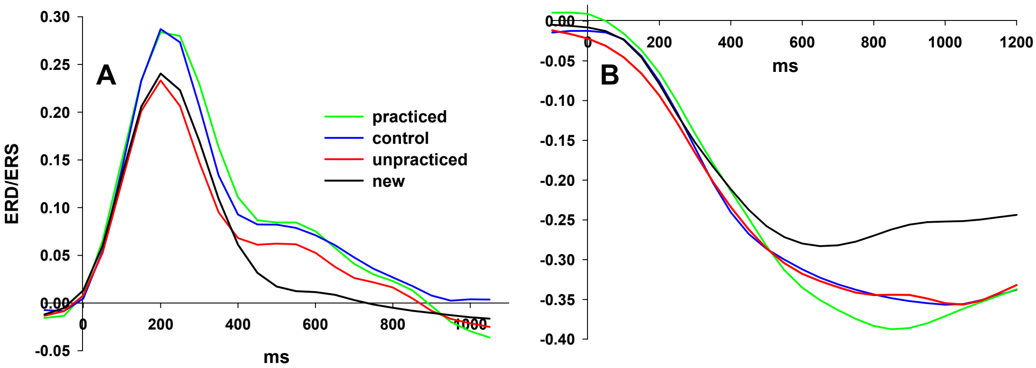

Figure 6: ERD/ERS pooled over parietal and frontal recording sites for each item type, Experiment 5. A: Theta (4-7 Hz); B: Upper Alpha (10-13 Hz). New: Correct rejections; Unpracticed: Hits to unpracticed items; Control: Hits to control items; Practiced: Hits to Practiced items.

P-, new), ROI (frontal vs parietal), and time window (early vs late) revealed no interaction between item type and ROI [$F(3,72) < 1$]. Therefore, the theta ERS data were collapsed across frontal and parietal recording sites for further analysis (Fz, F1, F2, F3, F4, FCz, FC1, FC2, FC3, FC4, Pz, P1, P2, P3, CPz, CP1, CP2, CP3).

A 4 × 2 analysis of variance with the two factors of item type (P+, C, P-, new) and time window (early vs late) revealed a significant main effect of item type [$F(3,72) = 4.48, MSE = 0.010, p < .01$] and a significant interaction between the two factors [$F(3,72) = 3.61, MSE = 0.002, p < .05$]. Subordinate analysis shows reliable effects of item type in the early time window [$F(3,72) = 3.73, MSE = 0.07, p < .05$] and in the late time window [$F(3,72) = 8.64, MSE = 0.004, p < .01$], indicating reliable theta old/new effects in both time windows. Planned comparisons revealed a reliably smaller theta ERS for unpracticed (P-) compared to control (C) items in the early time window [$t(24) = 2.11, p < .05$], but not in the later time window [$t(24) = 1.14, p > .25$]. There was no reliable increase in theta ERS for practiced (P+) compared to control items (C) [both $t(24) < 1$].

Upper Alpha ERD Results

The pooled upper alpha ERD results are shown in Figure 6B. Extensive screening of the data across various ROIs showed no reliable effects of topography on the degree of upper alpha ERD. Therefore, the upper alpha ERD data were collapsed across frontal and parietal recording sites for further analysis (Fz, F1, F2, F3, F4, FCz, FC1, FC2, FC3, FC4, Pz, P1, P2, P3, P4, CPz, CP1, CP2, CP3, CP4).

Visual inspection of the upper alpha ERD shows long-lasting old/new effects starting at about 600 ms after stimulus onset and reaching a maximum at about 900 ms. The ERD data were thus analyzed for a relatively late time window (750-1100 ms). A single-factor analysis of variance revealed a significant effect of item type [$F(3, 72) = 18.84, MSE = 0.004, p < .001$]. Planned comparisons revealed significant old/new effects for all three types of hits (P+, C, P-) [all $t(24)$'s> 4.93, all p's$< .001$]. Further, the old/new effect was marginally larger for P+ compared to C hits [$t(24) = 1.93, p < .10$], but not smaller for P- hits [$t(24) < 1$].

Theta Phase Coupling Results

Analysis of pairwise inter-electrode phase coupling as measured by the phase locking value (PLV) revealed no statistically reliable PLV-differences across item type in any of the frequency bands and time windows of interest (all $p > .05$).

Electrophysiological Results B: Remember/Know

For remember/know-analysis, the data from each subject were pooled across retrieval status (P+, C, P-).

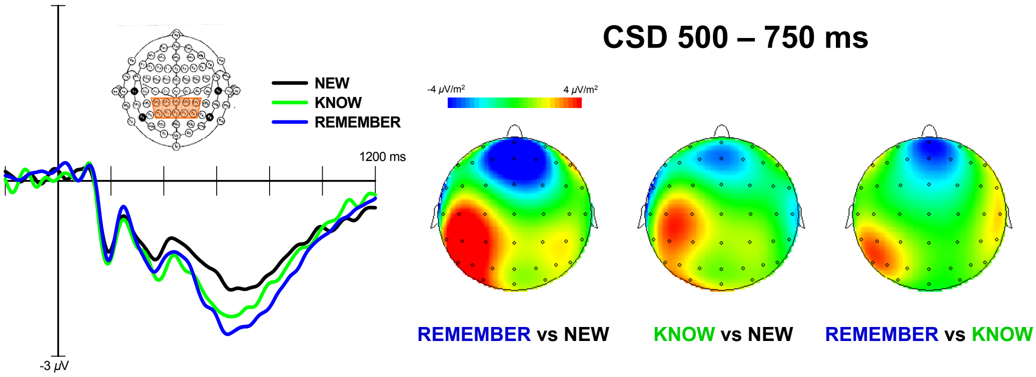

Figure 7: Remember/Know ERPs pooled over parietal recording sites and current scalp densities (CSD), Experiment 5. New: Correct rejections; Know: Know hits; Remember: Remember hits

ERP Results

The ERPs from parietal recording sites for remember-, know-, and new responses are depicted in Figure 7. Visual inspection of current scalp density maps indicates comparable topographical distributions for all three pairwise contrasts in the late time window.

A $3 \times 2 \times 2$ analysis of variance with the three factors of response category (remember, know, new), location (frontal vs parietal), and time window (early vs late) revealed a significant triple interaction between the three factors $[F(2, 48) = 6.10, MSE = 0.151, p < .01]$. Subordinate analyses showed no effects of response category and thus no significant old/new effects at frontal recording sites in the early time window $[F(2, 48) < 1]$. However, significant effects of response category emerged at parietal recording sites in the late time window $[F(2, 48) = 10.38, MSE = 0.263, p < .001]$. Planned comparisons revealed significant late old/new-effects at parietal recording sites for both know $[t(24) = 2.50, p < .05]$ and remember responses $[t(24) = 4.97, p < 0.001]$. This late parietal old/new effect was marginally

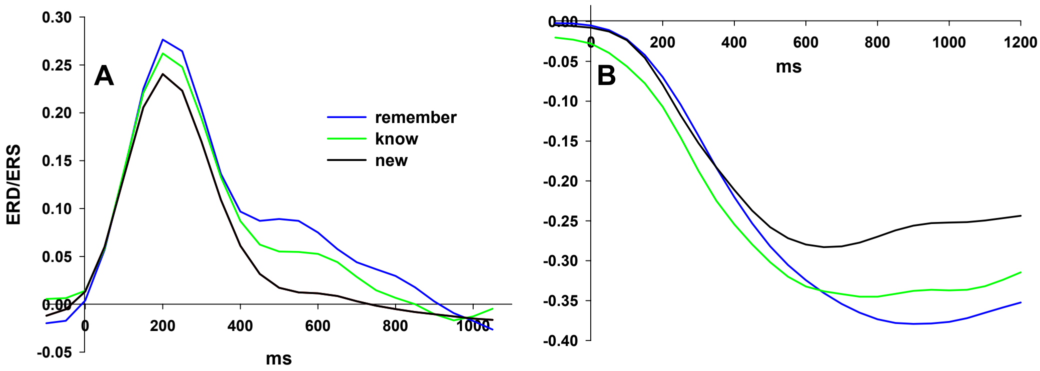

Figure 8: Remember/Know ERD/ERS, Experiment 5. A: Theta (4-7 Hz); B: Upper Alpha (10-13 Hz). New: Correct rejections; Know: Know hits; Remember: Remember hits

larger for remember than for know responses $[t(24) = 1.85, p < .10]$.

Theta ERS Results

The pooled theta ERS data for each item type are shown in Figure 8A. A 2 × 2 × 3 analysis of variance with the three factors of response category (remember, know, new), ROI (frontal vs parietal), and time window (early vs late), revealed no interaction between response category and ROI $[F(2, 48) < 1]$. Therefore, the theta ERS data were collapsed across frontal and parietal recording sites for further analysis (Fz, F1, F2, F3, FCz, FC1, FC2, FC3, FC4, Pz, P1, P2, P3, CPz, CP1, CP2, CP3).

A 3 × 2 analysis of variance with the two factors of response category (remember, know, new) and time window (early vs late) revealed a significant main effect of item type $[F(2, 48) = 4.93, MSE = 0.005, p < .05]$ but no significant interaction between the two factors $[F(2, 48) = 1.71, MSE = 0.003, p > .15]$, indicating significant theta ERS old/new effects in both time windows. Subordinate analyses shows no significant difference in overall theta ERS between remember- and know responses [$F(1,24) = 2.32, MSE = 0.004, p > .15$] and no interaction with factor time [$F(1,24) < 1$]. However, planned comparisons revealed a marginally larger theta ERS for remember compared to know hits in the late time window [$t(24) = 1.92, p < .10$].

Upper Alpha ERD Results

The pooled upper alpha ERD results are shown in Figure 8B. The remember/know ERD data were collapsed across the same frontal and parietal recording sites and analyzed in the same time window (750-1100 ms) as in the retrieval-induced forgetting analysis.

A single-factor analysis of variance revealed a significant effect of response category [$F(2,48) = 26.61, MSE = 0.004, p < .001$]. Planned comparisons revealed significant old/new effects for both response categories (remember, know) [both $t(24) > 5.20$, both $p < .001$]. Further, the upper alpha ERD was marginally larger for remember compared to know hits [$t(24) = 1.93, p < .10$].

Theta Phase Coupling Results

The phase locking value (PLV) results for the contrasts of interest are illustrated in Figure 9.

Compared to correct new responses, remember hits were associated with significantly stronger phase coupling between 500 and 800 ms after stimulus onset (95 electrode pairs, $p < .005$). The significant electrode pairs covered almost the whole scalp, with connections between temporal, parietal and occipital recording sites being particularly frequent. In contrast, no reliable old/new difference in phase locking was evident for know-hits (14 pairs, $p > .10$). In direct comparison with know hits, remember hits were associated with significantly enhanced fronto-parietal, fronto-temporal, and parieto-temporal phase coupling (37 pairs, $p < .05$).

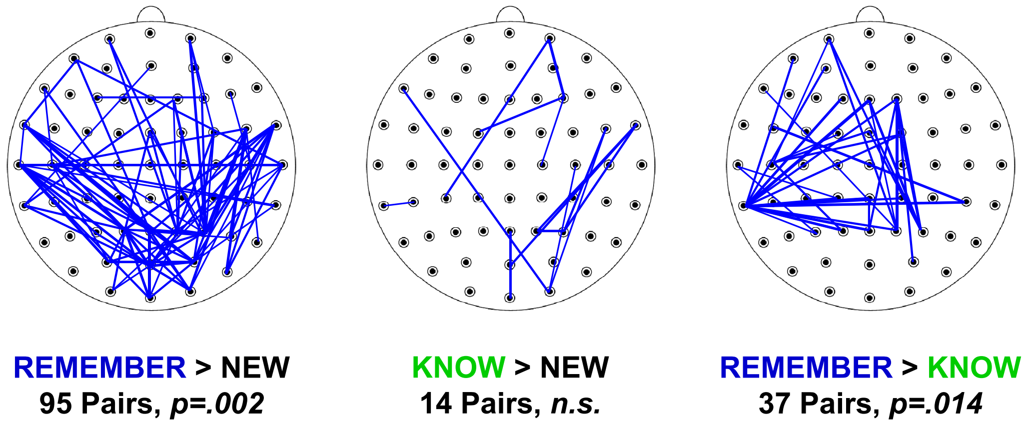

Figure 9: Remember/Know PLV, Experiment 5. New: Correct rejections; Know: Know hits; Remember: Remember hits. Blue lines indicate significantly stronger phase coupling for a given electrode pair ($p < .005$), line thickness indicates strength of phase coupling. Permutation tests were used to determine p-levels for each pattern.

Discussion

Retrieval-Induced Forgetting

Like in Experiments 1 and 2, using dual-process analysis methods, the behavioral results from Experiment 5 indicate a reliable decrease in the unpracticed items' familiarity. Further, the present experiment's dual-process estimates suggest a significant decline in the unpracticed material's recollection, whereas only nonsignificant trends for such a decline were evident in the foregoing experiments. However, this difference in finding might be attributable to differences in statistical power across experiments. In all, the present dual-process results are largely in accordance with the results from the previous experiments, suggesting that retrieval practice affects familiarity and - less robust across experiments - also recollection.

The overall pattern of ERP and ERD/ERS old/new effects across retrieval status is summarized in Figure 10. By tendency, each of the ERP

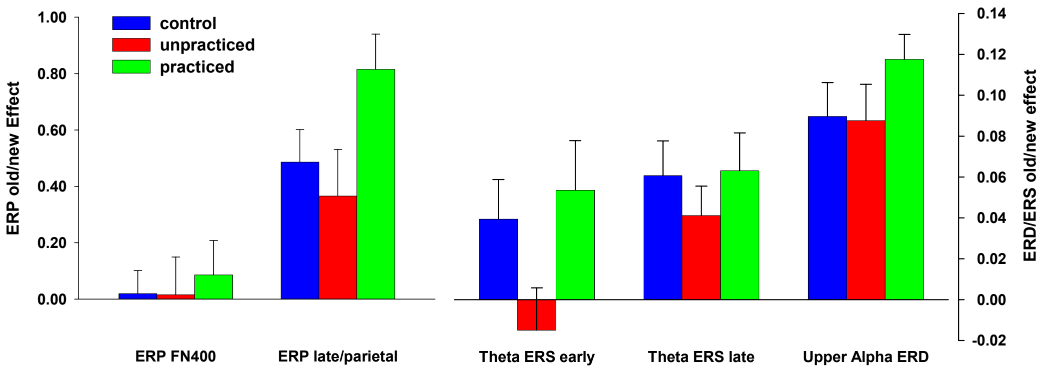

Figure 10: Summary of EEG old/new effects for each item type, Experiment 5.

and ERD/ERS measures exhibited the standard pattern of retrieval-induced forgetting, with larger old/new effects for practiced material, and smaller old/new effects for unpracticed material compared to control. However, the detrimental effects of retrieval practice were significantly reflected in early theta ERS only. In fact, the early theta ERS old/new effect appeared to be effectively eliminated for unpracticed items. Theta ERS during recognition of studied items is assumed to reflect the strength of the to-be-recognized items' episodic memory traces (see Klimesch et al., 2006). Moreover, early (200-400 ms) theta ERS has been tentatively associated with familiarity processes (see Klimesch et al., 2001; 2006). The present finding of a reduction in early theta ERS therefore agrees with the results from the foregoing behavioral experiments: Retrieval practice essentially reduces the familiarity (dual-process view) or general memory strength (single-process view) of semantically related unpracticed material.

Whereas the inhibitory mechanisms underlying retrieval-induced forgetting are assumed to directly affect an item's inherent memory representation, the strengthening of repeatedly processed material is typically attributed to an increase in binding between the material and its cue(s) (e.g., Raajmakers & Shiffrin, 1981; Rundus, 1973). In particular, in the present experiment, the retrieval-practice phase allowed for binding of the to-be-practiced material to a novel episodic context in addition to initial study. Consistently, recognition of practiced items was characterized by a significantly stronger late parietal ERP old/new effect. Indeed, a large body of literature has linked late parietal ERP old/new effects with episodic recollection of spatio/temporal information associated with the prior occurrence of a recognized stimulus (for reviews, see Curran et al., 2006; Friedman & Johnson, 2000 or Rugg & Yonelinas, 2003). The observed increase in parietal ERP positivity may thus reflect particularly rich episodic memories that are associated with the practiced material.

In Experiment 5, the detrimental and the beneficial effects of retrieval practice were qualitatively dissociable: The recognition correlate reflecting the weakening of the unpracticed material's memory traces (i.e., theta power) was not increased for practiced material and none of the correlates reflecting the strengthening of practiced material (late parietal positivity, upper alpha ERD) was reduced for unpracticed material. A priori, one might have expected that any recognition correlate which is reduced for unpracticed items (i.e., theta power) should be enhanced for practiced items, assuming that retrieval practice weakens the unpracticed material but at the same time strengthens the practiced material. However, a number of behavioral studies examining the effects of retrieval practice reported detrimental effects on the unpracticed material without enhancing effects on the practiced material (e.g., Gomez-Ariza et al., 2005; Veling & van Knippenberg, 2004) and beneficial effects on the practiced material without detrimental effects on the unpracticed material (e.g., Anderson, Green, & McCulloch, 2000; Bäuml & Kuhbandner, 2007). These findings indicate that the beneficial and detrimental effects of retrieval practice are mediated by different mechanisms, which is consistent with the present EEG analysis. Indeed, the EEG data suggest that the forgetting in this paradigm is due to a weakening of the item representation, whereas the behavioral enhancement reflects particularly rich episodic information associated with the practiced material.

A dual-process interpretation of the present EEG results may suggest a strong increase in recollection - but not familiarity - of the practiced material, as well as a reliable decrease in the unpracticed items' familiarity accompanied by only a small and unreliable decrease in recollection. However, the overall pattern of results points to several inconsistencies in the dual-process interpretation of the behavioral, ERP and theta ERS data. While the early theta ERS old/new effects - like the behavioral results - suggest a substantial role of familiarity in item recognition, no FN400 ERP old/new effects were observed (possible explanations will be considered in the next section). Further, whereas the late/parietal ERP old/new effect was significantly enhanced for practiced material, no such effect was evident in late theta ERS, although both types of old/new effects are assumed to reflect recollection (e.g., Klimesch et. al, 2001; Jacobs et al., 2006; Rugg & Yonelinas, 2003). Finally, the behavioral results indicate a strong increase in the practiced material's familiarity, but no such effect was evident in the putative theta ERS correlate of familiarity. Indeed, in the light of such considerable inconsistencies across analysis methods, a coherent dual-process interpretation of the present behavioral and electrophysiological results appears difficult.

On the basis of Experiments 1-4, it was suggested that the effects of retrieval practice may be best described as a modulation of general memory strength as conceptualized in single-process theories of recognition memory. It might be argued that such a single-process view of retrieval-induced forgetting implies that practiced, unpracticed and control items vary along a common theoretical dimension (i.e., general memory strength), ostensibly conflicting with the present finding of dissociable neural correlates of forgetting and facilitation. However, the 'single' process view does not exclude the possibility that modulations in general memory strength may be produced by a variety of mechanisms (see Background section) with potentially distinct neural signatures. Assuming that on the behavioral surface, both the strength of episodic memory traces and varying degrees of associated episodic information may in combination create continuous 'general memory

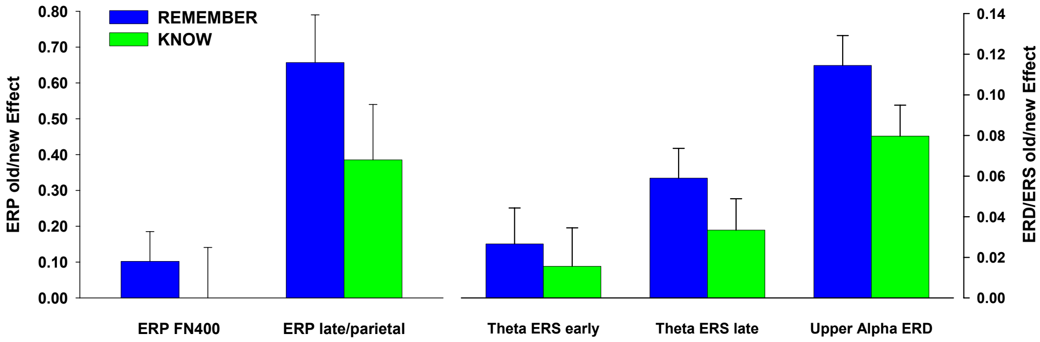

Figure 11: Summary of EEG-old/new effects for each response category, Experiment 5.

strength', the present EEG results can be plausibly integrated with a signal detection account of the behavioral data.

To conclude, the detrimental effects of inhibition on unpracticed items' recognition were characterized by reduced theta power, consistent with the view that retrieval-practice reduces the unpracticed material's general memory strength. The results also suggest that the detrimental effects and the beneficial effects of retrieval practice in item recognition are reflected by qualitatively dissociable EEG correlates, which is again consistent with prior behavioral work, and suggests that the beneficial and the detrimental effects of retrieval practice are mediated by different mechanisms.

Putative Correlates of Remembering and Knowing

The ERP and ERD/ERS old/new effects for remember and know responses are summarized in Figure 11. Although the behavioral results suggest that recollection and familiarity contributed about equally to overall recognition performance (see previous section), no FN400 ERP old/new effects emerged in the present experiment, probably due to particular methodological requirements of the retrieval-practice paradigm: Only relatively strong category exemplars are susceptible to retrieval-induced forgetting (Anderson, Bjork, & Bjork, 1994; Bäuml, 1998). Therefore, the study material used in the present experiment consisted of high-typicality words which most likely were pre-experimentally highly familiar to the subjects. It has been shown that the N400 ERP component is generally sensitive to semantic factors like typicality and word frequency (e.g. Kutas & Federmeier, 2000). The requisite use of highly typical verbal material might thus have generally alleviated (F)N400 old/new effects in the present experiment.

Due to the failure to elicit FN400 old/new effects, the present analysis may not be suited for an exhaustive discussion of the ERP correlates of remember- and know responses. Still, the present ERP results support the view of a quantitative rather than a qualitative distinction between remembering and knowing: Both types of responses were associated with reliable parietal/late ERP old/new effects that were topographically highly similar and only marginally enhanced for "remembered" items (Figure 7, for similar results, see Smith, 1993 or Trott et al., 1999). Such an interpretation of the present data is consistent with recent meta-analyses and reviews (Dunn, 2004; Wixted & Stretch, 2004) suggesting that remember judgments merely reflect relatively high levels of response confidence rather than specifically indicating the use of recollection (see also Donaldson, 1996; Wixted & Stretch, 2004).

Such an alternative interpretation of remember/know responses is also supported by the present finding that early and late ERS/ERD old/new effects yielded no dissociation of remembering and knowing. This finding conflicts with the view proposed by Klimesch et al. (2001; 2006) that early theta ERS (200-400 ms) reflects familiarity whereas prolonged theta ERS ($> 500ms$) reflects recollection. To date, however, the mapping of familiarity and recollection on different theta ERS time windows has been inferred from the results of a single study only (Klimesch et al., 2001) which employed the highly disputed remember/know procedure to separate memory processes. It appears possible that the present failure to replicate Klimesch et al.s' (2001) results simply reflects the inadequacy of the remember/know procedure to differentially assess recollection and familiarity processes (e.g., Dunn, 2004; Rotello et al., 2005).

Further EEG Correlates of Recognition Memory

Late Upper Alpha Desynchronisation

Beyond the well-known old/new effects in ERP and theta ERS reported above, explorative analysis of the present EEG data yielded old/new effects in the upper alpha frequency range (10-13 Hz). In a comparatively late time window (approx. 750-1100 ms), hits were generally characterized by significantly stronger upper alpha ERD compared to correct rejections. To date, only relatively few studies have reported late upper alpha old/new effects in episodic item recognition (e.g., Burgess & Gruzelier, 2000). However, frontal ERP old/new effects have often been observed for similarly late time windows, albeit not in the present study. Such late frontal ERP old/new effects have been tentatively associated with post-retrieval evaluation processes (for reviews, see Rugg & Allan, 2000, or Wilding & Sharpe, 2003), whereas event-related desynchronisation of alpha oscillations has been associated with active cognitive processing and complex spreading activation processes (for a review, see Klimesch et al., 2006b). It is suggested that the upper alpha old/new effect reflects monitoring processes, which may be related to the repeated processing of the material in prior phases of the experiment. This interpretation is consistent with the finding that the upper alpha ERD was selectively increased for practiced material and the increase was positively correlated with performance in the retrieval-practice phase[5]. Further work is needed to examine the issue in more detail.

Theta Phase Coupling

While ERP and ERD/ERS old/new effects yielded no dissociation of re-

[5] Across participants, the increase in ERD for practiced (P+) compared to control material in the test phase was positively correlated with performance in the prior retrieval-practice phase [MF: $r_{sp} = .52, p < .01$; LP: $r_{sp} = .68, p < .001$]

member and know hits, the PLV analysis of the present data suggests that inter-electrode synchronisation in the theta frequency range may be specific to 'remembering: Compared to new items, remember hits were associated with a strong increase in widespread theta phase coupling, while no such increase was evident for know hits. To the best of my knowledge, this is the first demonstration of a recognition correlate in theta PLV. In direct comparison to know hits, remember hits were associated with a significant increase in fronto-parietal, fronto-temporal, and parieto-temporal phase coupling (see Figure 9). Notably, the pattern of remember-specific phase coupling comprised recording sites that are typically associated with episodic recollection (parietal/temporal) and familiarity (frontal) in the dual-process literature (e.g Rugg & Yonelinas, 2003).

Inter-electrode synchronisation of brain oscillations is believed to reflect enhanced interaction and information exchange between distant brain regions (e.g. Fries, 2005; Lachaux et al., 1999; Varela et al., 2001). More specifically, with respect to memory processes, enhanced theta phase coupling during encoding has been associated with successful associative binding in episodic memory (Summerfield & Mangels, 2005). In the present context, it may be hypothesized that the widespread increase in theta phase coupling for 'remembered' items reflects the associative binding of most diverse spatio-temporal information into vivid representations of episodic memories.

On the basis of the present remember/know analysis, it can not be clearly determined whether the PLV contrast between 'remembering' and 'knowing' reflects a genuinely (i.e., qualitatively) different pattern of phase coupling or a merely quantitative difference. Interestingly however, whereas each of the ERP and ERD/ERS old/new effects reported earlier appeared to be basically sensitive to both response confidence (remember/know) and manipulations of item strength (P+, C, P-), no reliable PLV differences arose across retrieval status (P+, C, P-). Suggested from this observation, it appears possible that the increase in theta PLV may in fact be specific to subjective reports of 'remembering'.

Part IV

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 the 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, Experiments 2-4 by using the ROC procedure, and Experiment 5 by using the remember/know procedure and examining the putative EEG correlates of recollection and familiarity.

The five 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 four of the five experiments the forgetting was accompanied by a reliable reduction in the unpracticed items' familiarity. Only in Experiment 3, using semantically unrelated material that could be grouped by features of the studied items' perceptual context, no decrement in the unpracticed material's familiarity arose (see Figure 4B). In all, however, the dual-process results from the present experiments suggest that familiarity plays an important role in retrieval-induced forgetting.

The present finding of significantly reduced familiarity for the unpracticed material agrees with prior work demonstrating retrieval-induced forgetting in memory tests which are assumed to be substantially familiarity-driven, like speeded recognition (Veling & van Knippenberg, 2004) or old/new recognition (Gómez-Ariza et al., 2005; Hicks & Starns,2004). In particular, it agrees with the inhibitory account of retrieval-induced forgetting according to which retrieval practice weakens the unpracticed material's memory representation (Anderson, 2003; Anderson & Spellman, 2005). It was hypothesized that such weakening should not only affect recollection, as was suggested from prior work (e.g., Verde, 2004), but also familiarity, given that familiarity is thought to reflect an item's general memory strength (see Background section). The present dual-process analyses confirm this expectation on the basis of a considerable amount of empirical data.

Regarding the role of recollection in retrieval-induced forgetting, dual-process analysis yielded an unexpected pattern of results: Although, with the exception of Experiment 4, recollection appeared to substantially contribute to recognition performance, only nonreliable trends for a reduction in recollection of the unpracticed material arose in Experiments 1-3 (see Figure 4A). Because a large number of previous studies demonstrated retrieval-induced forgetting in free and cued recall tests which are assumed to be essentially recollection-driven, it was expected that recollection should also be impaired in the unpracticed material's recognition (see Background section). However, it can be hypothesized that the failure to detect significant reductions of recollection in Experiments 1-3 might have been due to insufficient statistical power. In accord with this conjecture, a statistically reliable reduction in the unpracticed material's recollection was evident in the behavioral data from Experiment 5, where EEG analysis required larger trial numbers than are typically used in the standard retrieval-practice paradigm. Thus, overall, it may be concluded that dual-process analysis does indicate a clear tendency for decreases in the unpracticed material's recollection, albeit these decreases appear to be generally smaller in size and less robust across experiments than the pronounced reductions in the unpracticed material's familiarity.

In three of the four experiments which employed tests of item recognition (Experiments 1, 2, and 5), both the forgetting of unpracticed material and the facilitation of practiced material were substantially familiarity-driven. In contrast, in Experiment 3, both forgetting and facilitation were exclusively recollection-driven. The main difference in methodology between these experiments was that in Experiment 3, episodically related material was used, whereas in the remaining three experiments, the study material was semantically related. The results may therefore suggest that retrieval practice on semantically related material modulates both recollection and familiarity for the practiced and the unpracticed items, whereas retrieval practice on episodically related material modulates recollection only (but see Experiment 3 for a critical discussion of this finding).

However, from a dual-process perspective, the specific effects of retrieval practice on recognition memory appear to depend not only on the structure of the study material, but also on the type of test: Suggested from prior work, both forgetting and facilitation are assumed to be primarily recollection-driven in associative recognition (Verde, 2004). In contrast, the present experiments suggest major effects on the material's familiarity in item recognition - at least with semantically related material - as well as in episodic feature recognition (see Experiment 4). Overall, it appears to depend in a complex manner on a variety of methodological variables whether or not retrieval practice modulates the affected material's recollection and/or familiarity in a given experiment. Due to such irregularities, the results from the present analyses can not be easily integrated in a coherent dual-process account of retrieval-induced forgetting.

To summarize, dual-process analysis of the present experiments yields strong evidence that retrieval-induced forgetting can substantially reduce the unpracticed material's familiarity. By comparison, the reductions in recollection appeared to be generally smaller and less robust across experiments. It may thus be concluded that by tendency, retrieval-induced forgetting in recognition is more an effect of familiarity than of recollection. However, the predictive value of such a general dual-process account of retrieval-induced forgetting is clearly limited given that the specific effects on recollection and familiarity may critically depend on a number of methodological factors.

Single-Process Account of the Data

The behavioral data of the present experiments were excellently described by the unequal-variance variant of the single-process signal detection model (e.g., Wixted, 2007; Wixted & Stretch, 2004). Notably, across experiments, the model fitted the data generally better than the dual-process model. These findings support a single-process view, according to which recognition performance 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 and increased the general memory strength of the practiced items (see Figure 4C). This held for all five experiments, using semantically and episodically related study material, employing tests of item and feature recognition, applying the remember/know procedure and the ROC procedure. The single-process account of the behavioral data thus provides a coherent explanation of retrieval-induced forgetting in the present study.

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 et al., 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.

To conclude, single-process analysis indicates that in each of the five experiments, retrieval practice essentially modulated the general memory strength of the practiced and the practiced material (see Figure 4C). In particular, the single-process models provide a compellingly parsimonious account of the present data that integrates seamlessly with prior work on retrieval-induced forgetting. Note that the present study was not designed to test the general validity of opposing models of recognition memory. In the literature, the question whether recognition performance is generally better described by a single- or by a dual-process model is a yet unresolved, fiercely disputed topic (e.g., Dunn, 2004; Parks & Yonelinas, 2007; Wixted 2007; Wixted & Stretch, 2004; Yonelinas, 2002). Regarding the competing models' accounts of retrieval-induced forgetting however, the present analyses indicate that the single-process account outperforms the dual-process account in terms of goodness of fit, theoretical parsimony, and predictive power.

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 experiments unpracticed items showed mainly deficits in familiarity but hardly any reduction 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. I 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 experiment, the two dual-process analysis methods converged on showing a stronger decline in recollection (specific memory strength) than in familiarity (global memory strength),[6] which is consistent with the descriptive analysis of the data reported by Verde himself. On the basis of the dual-process view, the present study and Verde's study thus lead to conflictive conclusions regarding the qualitative effects on unpracticed material's recollection and familiarity.

When fitting the single-process signal detection model to Verde's 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-

[6]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 to 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); none of the two effects was reliable, however [both $\chi^2(1)$'s $< 2.90, p > .05$]. 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 of unpracticed pairs compared to 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 a reliable decrease in d' (0.50 vs. 0.71) [$\chi^2(1) = 4.95$, $p < .05$].

process interpretation of the data is more parsimonious than the dual-process interpretation. In particular, it suggests a simple solution of the ostensible discrepancy between Verde's and the present findings. To resolve this discrepancy, the dual-process view has to assume that retrieval practice has qualitatively different effects in different recognition tasks (i.e., item recognition, inter-item associative recognition, and associative feature recognition). In contrast, the single-process view can account for the results in the various types of tasks on a purely quantitative basis, i.e., by assuming that the amount of reduction in unpracticed items' general memory strength can vary across type of task.

Effects of Feature Binding

Ciranni and Shimamura (1999) reported retrieval-induced forgetting of nonverbal stimuli that shared perceptual features with the practiced material. In Experiment 3, using semantically unrelated verbal stimuli that could be grouped by context color, retrieval of a subset of the words impaired later item recognition of episodically related words, thus generalizing previous findings of retrieval-induced forgetting in episodic memory. However, whereas in Ciranni & Shimamura's (1999) study the forgetting generalized to various perceptual features of the unpracticed material, memory for the unpracticed items' episodic contexts was unaffected in Experiment 3. Thus, on the one hand, the present findings support the view that retrieval-induced forgetting is a general phenomenon that can be observed across a large variety of experimental settings and testing procedures. On the other, the results also reveal possible limitations of such forgetting. On the basis of Experiment 3, it may be suggested that episodic context information is not susceptible to retrieval-induced forgetting.

The results from Experiment 4 further elucidate such limitations of retrieval-induced forgetting in episodic memory. In Experiment 4, subjects studied semantically unrelated words that could be grouped by their font color. Interestingly, retrieval of a subset of the *words* had a detrimental effect on recognition of the unpracticed words' *colors*. Thus, unlike Experiment 3, Experiment 4 replicates Ciranni and Shimamura's (1999) finding that the forgetting can generalize to perceptual item features that had not been actively retrieval practiced.

Critically, in Experiment 4, like in Ciranni and Shimamura's (1999) study, the episodic information 'color' was presented as an integral feature of the item's appearance, whereas in Experiment 3, the color information was a feature of the material's perceptual context. The differences in finding across experiments may thus be plausibly accounted for by feature binding mechanisms: Research on feature binding in episodic memory suggests that an object's color is easily *bound* into a single memory representation of a colored object, whereas presenting items against colored backgrounds rather creates associative links between the objects' representations and the respective background colors (e.g., Ecker et al., 2004; Groh-Bordin et al., 2006). Assuming that retrieval inhibition directly weakens the interfering items' memory representations (e.g., Anderson, 2003), retrieval-induced forgetting should mainly affect episodic information that is actually part of (i.e., 'bound' to) these memory representations. Consistent with such a proposal, the present findings of impaired memory for *item* color (Experiment 4) and unaffected memory for *context* color (Experiment 3) indicate that 'unbound' episodic context information may in fact not be susceptible to retrieval inhibition.

Proportional Strengthening predicts Changes in Target Variance

The present experiments' behavioral data were excellently described by the unequal-variance signal detection model, which suggests that the detrimental and the beneficial effects of retrieval practice can be entirely explained by unidimensional (i.e., quantitative) variations in general memory strength as measured by parameter d'. However, in two of the reported experiments, the strengthening of practiced material was further accompanied by reliable changes in parameter σ. Notably, in Experiment 3, σ was significantly larger for practiced compared to control items, indicating an *increase* in target variance, whereas in Experiment 5, σ was reliably smaller, suggesting a *decrease* in target variance. One might argue that such unsystematic modulations in σ may reflect *qualitative* differences in the practiced material's recognition, which conflicts with the view that recognition performance relied on a single source of mnemonic information.

However, detailed inspection of the parameter estimates derived from the unequal-variance signal detection model shows that the modulations in parameter σ were predictable on the basis of modulations in parameter d'. More specifically, across experiments, the modulations in practiced material's σ (expressed as σ_{P+}/σ_C) and the proportional increases in practiced material's d' (expressed as $(d'_{P+} - d'_C)/d'_C$) were almost perfectly correlated [$r = .96, p < .01$] (see Figure 12).[7]

Suggested from this analysis, the ostensibly unsystematic changes in the practiced material's target variance across experiments were predictable on the basis of proportional increases in the material's general memory strength. It can thus be assumed that the modulations in d' and the changes in σ reflect the assessment of a single source of mnemonic information which varies in strength only. This suggestion is in line with the single-process account of

[7]For the present demonstration, changes in σ were expressed as ratios of σ's for practiced to σ's for control items. Ratios > 1 indicate an increase in σ, ratios < 1 indicate a decrease. Expressing changes in σ as differences (i.e., $\Delta(\sigma)$) yielded an identical correlation. However, because σ is defined as the *ratio* of target- to noise variances, ratios of σ's were used. An identical correlation was also obtained when using alternative recognition performance measures (i.e., hit rate - false alarm rate). However, because general memory strength is modeled in z-space, d's were used for the present demonstration. The data from Experiment 4 were excluded from the present analysis because they could be satisfactorily described by an equal-variance model.

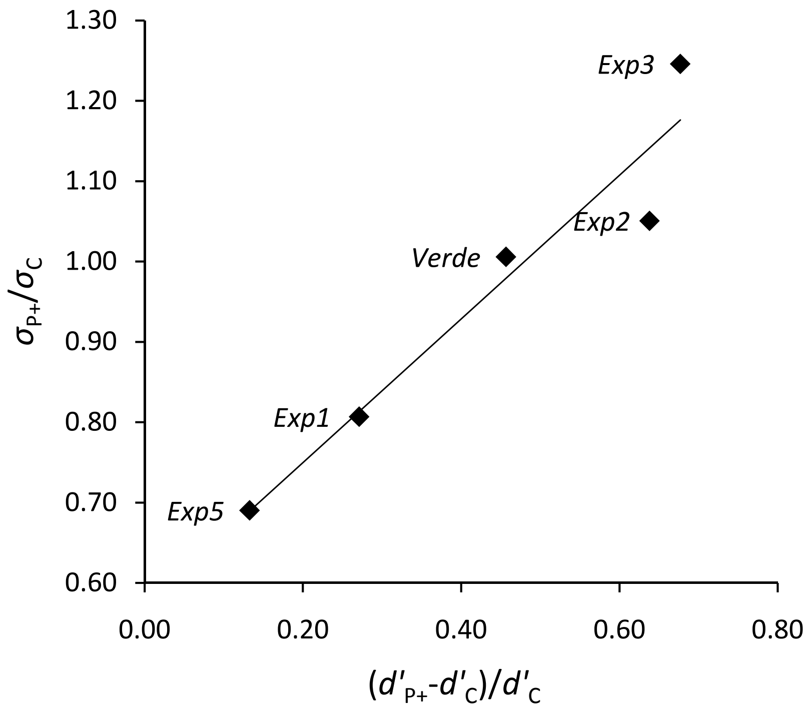

Figure 12: Linear relation between proportional increases in d' $(d'_{P+} - d'_C)/d'_C)$ and changes in σ for practiced compared to control material (σ_{P+}/σ_C) across experiments, including Verde's (2004) Experiment 1.

the present behavioral data, which implies that retrieval practice modulates recognition (i.e., performance) quantitatively but not qualitatively.

EEG Correlates of Retrieval-Induced Forgetting

Suggested from the analysis of recognition performance, retrieval-practice substantially reduces the general memory strength (single-process view) or familiarity (dual-process view) of the unpracticed material. Going beyond the analysis of behavioral memory performance, the EEG analysis of Experiment 5 indicates that recognition of unpracticed material is characterized by reduced theta ERS. In previous studies, it has been shown that the degree of such theta ERS is related to the strength of a to-be-recognized items' episodic memory trace (Klimesch et al., 2001; 2006). The present finding of a reduction in theta ERS therefore agrees with the view that retrieval practice reduces the unpracticed material's general memory strength or familiarity.

In contrast, recognition of practiced material was characterized by increased late parietal ERP old/new effects, which have been related to episodic recollection of spatio/temporal information associated with the prior occurrence of a recognized stimulus (for reviews, see Curran et al., 2006; Friedman & Johnson, 2000 or Rugg & Yonelinas, 2003). The overall pattern of results thus indicates that the detrimental effects and the beneficial effects of retrieval practice in item recognition are reflected by qualitatively dissociable EEG correlates, which is again consistent with prior behavioral work, and may reflect that the beneficial and the detrimental effects of retrieval practice are mediated by different mechanisms. More specifically, the results suggest that the behavioral enhancement in the retrieval-practice paradigm reflects particularly rich episodic information associated with the practiced material, whereas the forgetting is due to an inhibitory weakening of the item representation. It is argued that inhibited material elicits only relatively weak memory signals at test, reflected by selectively reduced theta power.

A dual-process interpretation of the present EEG results may suggest a strong increase in recollection - but not in familiarity - for the practiced material, as well as a reliable decrease in the unpracticed items' familiarity accompanied by only a small and unreliable decrease in recollection. However, the comparative analysis of behavioral dual-process estimates and putative EEG correlates of recollection and familiarity yielded partly incongruent results across measurement methods, which may point to internal inconsistencies of the neuro-cognititive dual-process framework (see next section). Alternatively, the present EEG results may be integrated with a signal-detection account of the behavioral data, assuming that various functionally distinct neural mechanisms may in combination create varying degrees of general memory strength.

EEG Correlates of Remembering and Knowing

The analysis of EEG correlates of 'remembering' and 'knowing' (Experiment 5) yielded no reliable dissociations of the putative EEG correlates of recollection and familiarity across response categories. Rather, the results supported a quantitative distinction between remember- and know responses. Consistent with prior research on the remember/know procedure (e.g., Dunn 2004; Wixted & Stretch, 2004), remember responses appeared to reflect generally stronger - but not different - memory processes compared to know responses.

The present pattern of results is inconsistent with EEG studies reporting dissociable EEG correlates of remembering and knowing (e.g., Düzel et al., 1997; Klimesch et al., 2001; Rugg et al.,1998). In part, the inconsistencies across studies may be attributable to insufficient sensitivity of (F)N400 old/new effects when high-frequency verbal stimuli are used in tests of item recognition (e.g., Kutas & Federmeier, 2000). On the other hand, a number of studies reported results which did not agree with the popular dual-process interpretation of ERP old/new effects (e.g., Trott et al., 1999; Finnigan et al., 2002; Voss and Paller, 2006; Yovel and Paller, 2004). However, the most plausible explanation for the pattern of results obtained in the present study might be that the remember/know procedure does in fact not allow for a process-pure dissociation of recollection and familiarity (e.g., Dunn, 2004; Rotello et al., 2005; Wixted & Stretch, 2004). Such an explanation may also account for the observed discrepancies in finding across behavioral and electrophysiological dual-process measures (see Experiment 5, Retrieval-Induced Forgetting Analysis).

In the present study, remember judgements were accompanied by a specific increase in inter-electrode synchronisation in the theta frequency range (500-800 ms after stimulus onset), comprising frontal, parietal, and temporal recording sites. On the basis of prior work (Summerfield & Mangels, 2005), it is suggested that such widespread theta phase coupling in item recognition may reflect the associative binding of spatio/temporal information into vivid episodic memories. Taking further into account previous critique of the remember/know procedure (e.g., Dunn, 2004; Rotello et al., 2005; Wixted & Stretch 2004; Wixted, 2007), it appears possible that such binding processes may underlie the subjective awareness of 'recollecting something' rather than promoting a recollection process as incorporated in dual-process models of recognition memory.

Such a proposal is consistent with the results from a single case study by Barbeau et al. (2005) that reported a relation between inter-electrode synchronisation in the theta band and subjective experiences of recollection. The authors successfully elicited illusory recollective experiences in an epileptic subject after intracranial stimulation of medial and temporal structures. The artificially induced hallucinations of recollection were associated with a significant increase in theta synchronisation between widespread intracranial recording sites. Importantly, the theta synchronisation found in Barbeau et al.'s (2005) study characterized merely the subjective feeling of remembering recollective information, unrelated to any actual memory performance at all.

The idea that remember responses reflect subjective awareness of recollection rather than 'recollection itself' is not new to memory research. In fact, the 'remember'-response was originally designed by Tulving (1985) to identify a special state of subjective awareness termed 'autonoetic awareness' (Tulving, 1983), which is often described as a 'mental time travel back to the original encoding event'. Critically, such autonoetic awareness is thought to be a possible result (i.e., by-product) of assessing particularly strong memories whereas in dual-process theory, recollection is assumed to operate independently, effectively contributing to recognition memory. Wixted (2007) proposed that autonoetic awareness may be a more threshold-like process, and that thus neural mechanisms which appear to be specific for remember responses may reflect autonoetic awareness rather than recollection. It appears possible that the theta PLV old/new effects found in the present EEG analysis may specifically reflect such states of consciousness awareness associated with remember responses.

Conclusions

Prior work showed that retrieval practice can affect unpracticed material's recollection. The present study complements these findings by demonstrating that retrieval-practice can also reduce the unpracticed material's familiarity (dual-process view) or general memory strength (single-process view), which is consistent with the inhibitory account of retrieval-induced forgetting. Recognition of unpracticed material was further characterized by reduced theta activity in the EEG, indicating that inhibited material triggers only relatively weak memory signals. The findings are consistent with the view that retrieval inhibition leads to persistent weakening of the affected material's memory representation.

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.

Barbeau, E., Wendling, F., Régis, J., Duncan, R., Poncet, M., Chauvel, P., & Bartolomei, F. (2005). Recollection of vivid memories after perirhinal region stimulations: synchronization in the theta range of spatially distributed brain areas *Neuropsychologia*, *43*, 1329-1337.

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. (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*, 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.

Burgess, A. P. & Gruzelier, J. H. (2000). Short duration power changes in the EEG during recognition memory for words and faces. *Psychophysiology*, *37*, 596-606.

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

Curran, T. (2000). Brain potentials of recollection and familiarity. *Memory & Cognition, 28*, 923-938.

Curran, T. (2004). Effects of attention and confidence on the hypothesized ERP correlates of recollection and familiarity. *Neuropsychologia*, *42*, 1088-1106.

Curran, T. & Cleary, A. M. (2003). Using ERPs to dissociate recollection from familiarity in picture recognition. *Cognitive Brain Research*, *15*, 191-205.

Curran, T., Tepe, K. L., & Piatt, C. (2006). ERP explorations of dual processes in recognition memory. In H. D. Zimmer, A. Mecklinger & U. Lindenberger (Eds.) *Binding in Human Memory: A Neurocognitive Approach* (pp. 467-492). Oxford: Oxford University Press.

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

Donaldson, W. (1996). The role of decision processes in remembering and knowing. *Memory & Cognition*, *24*, 523-533.

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.

Düzel, E., Yonelinas, A. P.,Mangun, G. R.,Heinze, H.-J., & Tulving, E. (1997). Event-related potential correlates of two states of conscious awareness in memory. Proceedings of the *National Academy of Sciences of the United States of America*, *94*, 5973-5978.

Ecker, U. K. H., Groh-Bordin, C., & Zimmer, H. D. (2004). Electrophysiological correlates of specific feature binding in remembering - Introducing a neurocognitive model of human memory. In A. Mecklinger, H. D. Zimmer, & U. Lindenberger (Eds.) *Bound in Memory:*

Insights from Behavioral and Neuropsychological studies (pp. 159-193). Aachen: Shaker.

Eichenbaum, H. & Bunsey, M. (1995). On the binding of associations in memory: clues from studies on the role of the hippocampal region in paired-associate learning. *Current Directions in Psychological Science*, *4*, 19-23.

Finnigan, S., Humphreys, M.S., Dennis, S., & Geffen, G. (2002). ERP old/new effects: memory strength and decisional factor(s). *Neuropsychologia*, *40*, 2288-2304.

Fries, P. (2005). A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. *Trends in Cognitive Sciences*, *9*, 474-780.

Friedman, D. & Johnson, R., Jr (2000). Event-related potential (ERP) studies of memory encoding and retrieval: a selective review. *Microscopy Research and Technique*, *51*, 6-28.

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.

Gonsalves, B. D., Kahn, I., Curran, T., Norman, K. A., & Wagner, A. D. (2005). Memory strength and repetition suppression: Multimodal imaging of media temporal cortical contributions to recognition. *Neuron*, *47*, 751-761.

Groh-Bordin, C., Zimmer, H. D., & Ecker, U. K. H. (2006). Has the butcher on the bus dyed his hair? When color changes modulate ERP correlates of familiarity and recollection. *NeuroImage*, *32*, 1879-1890.

Hari, R. & Salmelin, R. (1994). Spatiotemporal characteristics of sensorimotor neuromagnetic rhythms related to thumb movement. *Neuroscience*, *60*, 537-550.

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.

Jacobs, J., Hwang, G., Curran, T., & Kahana, M. J. (2006). EEG oscillations and recognition memory: Theta correlates of memory retrieval and decision making *NeuroImage*, *32*, 978-987.

Jacoby, L. L. (1991). A process dissociation framework: Seperating automatic from intentional uses of memory. *Journal of Memory and Language, 30*, 513-541.

Jäger, T., Mecklinger, A., & Kipp, K. (2006). Intra- and inter-item associations doubly dissociate the electrophysiological correlates of familiarity and recollection. *Neuron*, *52*, 535-545.

Johansson, M., Aslan, A., Bäuml, K.-H., Gäbel, A., & Mecklinger, A. (in press). When remembering causes forgetting: electrophysiological correlates of retrieval-induced forgetting. *Cerebral Cortex*.

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

Klimesch, W., Doppelmayr, A., Yonelinas, A., Kroll, N. E. A., Lazzara, M., Röhm, D., & Gruber, W. (2001). Theta synchronization during episodic retrieval: Neural correlates of conscious awareness. *Cognitive Brain Research*, *12*, 33-38.

Klimesch, W., Doppelmayr, A., Stadler. W., Pöllhuber, D., Sauseng, P., & Röhm, D. (2000). Theta oscillations and the ERP old/new effect: independent phenomena?. *Clinical Neurophysiology*, *111*, 781-793.

Klimesch, W., Hanslmayr, S., Sauseng, P., Gruber, W., Brozinsky, C. J., Kroll, N. E. A., Yonelinas, A. P., & Doppelmayr, M. (2006). Oscillatory eeg correlates of episodic trace decay. *Cerebral Cortex*, *16*, 280-290.

Klimesch, W., Sauseng, P., & Hanslmayr, S. (2007). EEG alpha oscillations: The inhibition-timing hypothesis. *Brain Research Reviews*, *53*, 63-88.

Kutas, M. & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences*, *4*, 463-470.

Lachaux, J. P., Rodriguez, E., Martinerie, J., & Varela, F. J. (1999). Measuring phase synchrony in brain signals. *Human Brain Mapping*, *8*, 194-208.

MacKenzie, G. & Donaldson, D. I. (2007). Dissociating recollection from familiarity: Electrophysiological evidence that familiarity for faces is associated with a posterior old/new effect. *NeuroImage*, *36*, 454-463.

Macmillan, N. A. & Creelman, C. D. (2005). *Detection theory: A user's guide.* (2nd Ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

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.

Manns, J. R., Hopkins, R. O., Reed, J. M. , Kitchener E. G., & Squire, L. R. (2003). Recognition Memory and the Human Hippocampus *Neuron*, *37*, 171-180.

Mandler, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological Review, 87*, 252-271.

Mannhaupt, H.-R. (1983). Reproduktionsnormen für verbale Reaktionen zu 40 geläufigen Kategorien. *Sprache und Kognition, 4*, 264-278.

Parks, C. M. & Yonelinas, A. P. (2007). Moving beyond pure signal-detection models: Comment on Wixted (2007) *Psychological Review, 114*, 188-201.

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.

Pfurtscheller, G., 1992. Event-related synchronisation (ERS): an electrophysiological correlate of cortical areas at rest. *Electroencephalography and Clinical Neurophysiology. 83*, 62-69.

Pfurtscheller, G. & Aranibar, A. (1977). Event-related cortical desynchronization detected by power measurements of scalp EEG. *Electroencephalography and Clinical Neurophysiology. 42*, 817-826.

Quamme, J. R., Yonelinas, A. P., Widaman, K. F., Kroll, N. E. A., and Sauvé , M. J. (2004). Recall and recognition in mild hypoxia: Using covariance structural modeling to test competing theories of explicit memory. *Neuropsychologia*, *42*, 672-691.

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.

Rugg, M.D. and Allan, K. (2000). Event-related potential studies of memory. In Tulving E. & Craik, F. I. M. (Eds). *Oxford Handbook of Memory* (pp 521-537). New York: Oxford University Press.

Rugg, M. D. & Doyle, M. C. (1992). Event-related potentials and recognition memory for low- and high-frequency words. *Journal of Cognitive Neuroscience*, *5*, 69-79.

Rugg, M. D., Mark, R. E., Walla, P., Schloerscheidt, A. M., Birch, C. S., & Allan, K. (1998). Dissociation of the neural correlates of implicit and explicit memory. *Nature*, *392*, 595-598.

Rugg, M. D. & Yonelinas, A. P. (2003). Human recognition memory: A cognitive neuroscience perspective. *Trends in Cognitive Sciences, 7*, 313-319.

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

Saunders, J. & MacLeod, M. D. (2006). Can inhibition resolve retrieval competition through the control of spreading activation? *Memory & Cognition, 34*, 307-322.

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

Smith, M. E. (1993). Neurophysiological manifestations of recollective experience during recognition memory judgments. *Journal of Cognitive Neuroscience*, *5*, 1-13.

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.

Summerfield, C. & Mangels, J. A. (2005). Coherent theta-band EEG activity predicts item-context binding during encoding. *NeuroImage*, *24*,692-703.

Trott, C. T., Friedman, D., Ritter, W., Fabiani, M., & Snodgrass, J.G. (1999). Episodic priming and memory for temporal source: event-related potentials reveal age-related differences in prefrontal functioning. *Psychology and Aging*, *14*, 390-413.

Tulving, E. (1983). *Elements of episodic memory.* Oxford University Press.

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

Varela, F., Lachaux, J. P., Rodriguez, E., Martinerie, J. (2001). The brainweb: phase synchronization and large-scale integration. *Nature Neuroscience Reviews*, *2*, 229-239.

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.

Vilberg, K. L. & Rugg, M. D. (2007) Dissociation of the neural correlates of recognition memory according to familiarity, recollection, and amount of recollected information. *Neuropsychologia*, *45*, 2216-2225.

Voss, J. L., and Paller, K. A. (2006). Fluent conceptual processing and explicit memory for faces are electrophysiologically distinct. *Journal of Neuroscience*, *26*, 926-933.

Wilding, E. L. and Sharpe, H. (2003). Episodic memory encoding and retreival: Recent insights from event-related potentials. In Zani, A. & Proverbio A. M. (Eds) *The Cognitive Electrophysiology of Mind And Brain* (pp. 169-196). San Diego, CA: Academic Press.

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.

Woodruff, C. C., Hayama H. R., & Rugg, M. D. (2006). Electrophysiological dissociation of the neural correlates of recollection and familiarity *Brain Research*, *1100*, 125-135.

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. (1999). The contribution of recollection and familiarity to recognition and source-memory judgements: A formal dual-process analysis of reciever operating characteristics. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*, 1415-1434.

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.

Yonelinas, A. P., Otten, L. J., Shaw, K. N. & Rugg, M. D. (2005). Seperating the brain regions involved in recollection and familiarity in recognition memory. *Journal of Neuroscience*, 25, 3002-3008.

Yovel, G., & Paller, K. A. (2004). The neural basis of the butcher-on-the-bus phenomenon: When a face seems familiar but is not remembered. *Neuroimage*, *21*, 789-900.

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

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 (Figure 13A). Assuming equal variances of the underlying strength distributions, the probability of correctly recognizing a studied item (H) is given by

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

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

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

where d' is the distance between the means of the underlying distributions for new vs old items. Across different levels of response confidence (c_i), the equal-variance signal detection model theoretically yields a curvilinear ROC that is symmetrical along the diagonal (Figure 13C).

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 (Figure 13B). Therefore, the model's equal-variance restriction is

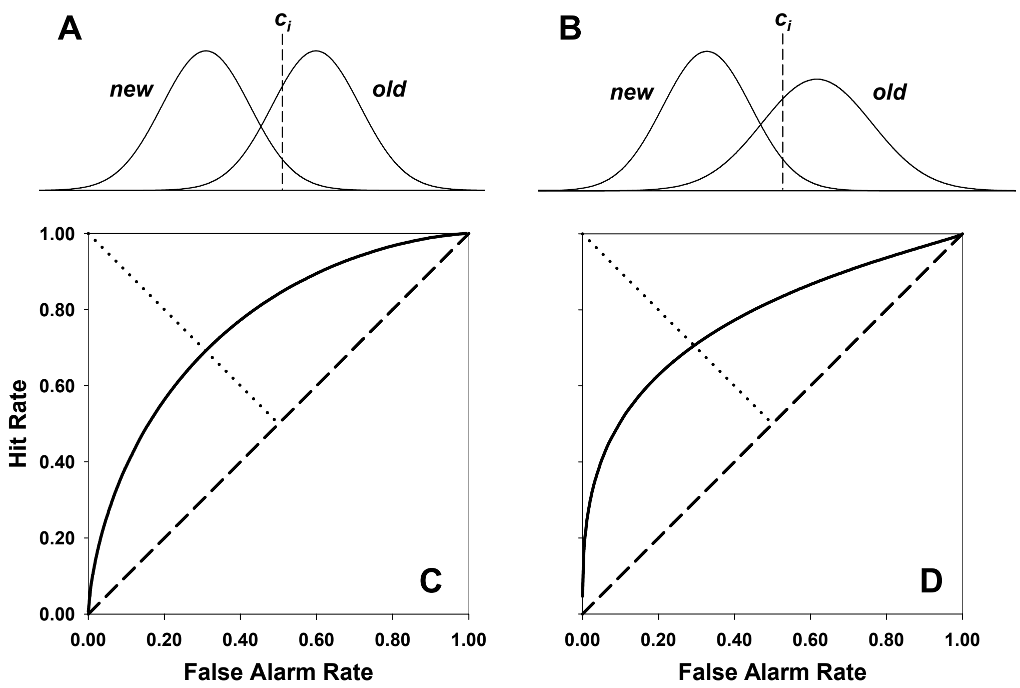

Figure 13: Theoretically assumed distributions of general memory strength and ROC-curves predicted by single process signal detection models of recognition memory. A: Equal-variance strength distributions; B: Unequal-variance strength distributions; C: ROCs predicted by the equal variance signal detection model; D: ROCs predicted by the equal variance signal detection model

usually relaxed by letting the variance of the target distribution vary freely, which leads to

$$p(H) = 1 - \Phi((c_i - d)/\sigma),$$

where σ is the variance of the old items' distribution given that the variance of the new items' distribution is set to 1. Figure 13D shows a typical theoretical ROC generated by the unequal-variance signal detection model.

For data from remember/know experiments, the equal-variance signal detection model has three free parameters (d' and criteria for remember and know) to account for four data points (hits and false alarms for remember and know). Thus, the models' goodness of fit can be statistically tested with one degree of freedom, whereas the unequal-variance model (with additional parameter s) is technically saturated and thus can not be tested with remember/know data. For ROC data derived from six-point confidence rating scales, the unequal-variance model has seven free parameters (d', σ, c_1 - c_5) to account for ten data points (hits and false alarms for confidence levels 1-5), as cumulative response frequencies naturally add to unity at confidence level 6). Thus, the model has three degrees of freedom to test its goodness of fit. using maximum-likelihood techniques. Accordingly, the maximum likelihood techniques described above for the dual-process signal detection model can be analogously applied to the unequal-variance signal detection model.

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 (Figure 14A), whereas familiarity can be described by a classical signal detection process as outlined above for the equal-variance

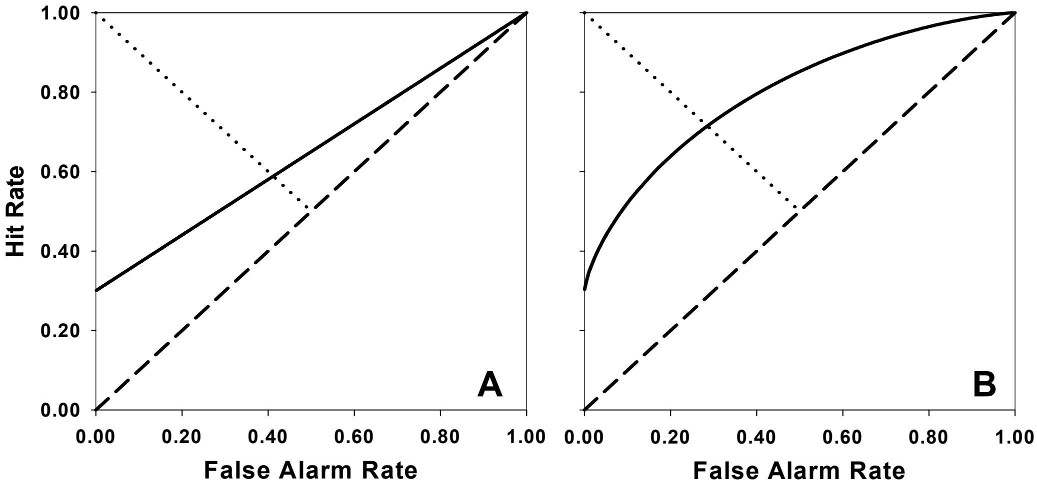

Figure 14: Theoretical ROC-curves predicted by the dual-process signal detection model (Yonelinas, 1994). A: Recollection only; B: Recollection and Familiarity

single-process model (Figures 13A/C). 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):

$$p(H) = R + (1 - R)(1 - \Phi(c_i - d')).$$

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

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

because it is usually assumed that new items can not be recollected. Figure 14B shows a typical theoretical ROC generated by the dual-process model. Except for extremely high-confidence regions of the ROC, the shape of the dual-process ROC is very similar to the ROC predicted by the unequal-variance model (Figure 13D). For ROC data derived from six-point confidence rating scales, the dual-process model has seven free parameters (d', R, c_1 - c_5) to account for ten data points (hits and false alarms for confidence levels 1-5). Thus, the model has three degrees of freedom to test its goodness of fit.

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, and $p(H)_i$ and $p(FA)_i$ are the model equations predicting these rates as functions of the model's parameters. 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 STREAK model. To test statistically whether certain model parameters vary significantly across item type, likelihood ratio tests were used. The logic behind likelihood ratio testing is to impose certain restrictions on the parameters of a given model and compare the restricted model's fit with the fit of the unrestricted model. For example, to test whether a parameter varies across two item types, the fit of a model with that parameter held constant is compared to the fit of the unrestricted model allowing that parameter to vary freely across item type. The difference in goodness of fit can be expressed in terms of a χ^2-distributed statistic with one degree of freedom (reflecting the number of restrictions imposed on the model's parameters) for statistical testing. If the statistic exceeds the critical χ^2-value, a significant change of the parameter across item type is indicated.