Category Cueing: Increasing Recall to Reduce False Recognition

Inaugural-Dissertation

zur

Erlangung des Doktorgrades der
Mathematisch-Naturwissenschaftlichen Fakultät
der Heinrich-Heine-Universität Düsseldorf

vorgelegt von

Juliane Schmid

aus Schorndorf

Mai 2008

Aus dem Institut für Experimentelle Psychologie

der Heinrich-Heine-Universität Düsseldorf

(Lehrstuhl Prof. Dr. A. Buchner)

Gedruckt mit der Genehmigung der
Mathematisch-Naturwissenschaftlichen Fakultät der
Heinrich-Heine-Universität Düsseldorf

Referent: Prof. Dr. A. Buchner

Korreferent: PD Dr. P. Jansen-Osmann

Tag der mündlichen Prüfung: 04.07.2008

Table of Contents

Abstract

The present work focuses on recall-to-reject, one of the central memory editing mechanisms thought to prevent the occurrence of false memories. Recall-to-reject occurs when recall of a studied item is used to correctly reject an associated distractor despite its high familiarity. In a series of four experiments, category cues were used to increase the probability of the recall-to-reject process. When semantically associated distractors were used in an item recognition task (Experiments 1 and 2), category cues eliminated the false recognition effect but the overall level of recognition errors was not reduced due to a simultaneous familiarity increase and participants' tendency to answer consistently. These side effects, however, could be successfully eliminated by using phonologically associated distractors: In Experiment 3, the suppression of the false recognition effect with category cues was replicated and, moreover, the false alarm rate was reduced. Furthermore, the results of Experiments 1 and 3, where no explicit recall-to-reject instructions were given, reveal that participants spontaneously use recall-to-reject without explicitly being instructed to. The results of Experiment 4, where associated items were included in the study list, reveal the important role of mutual exclusivity of stimuli for the effectiveness of recall-to-reject. Overall, the results of the reported experiments provide strong evidence for recall-to-reject and demonstrate its potential to reduce false recognition errors. Therefore, the results are consistent with dual-process theories of recognition memory.

Zusammenfassung

Die vorliegende Arbeit befasst sich mit dem *recall-to-reject* Prozess, einem der sogenannten *memory editing* Prozesse, von denen angenommen wird, dass sie das Auftreten falscher Erinnerungen verhindern können. Von *recall-to-reject* spricht man, wenn die bewusste Erinnerung an ein gelerntes Item zur korrekten Zurückweisung eines dazu assoziierten Distraktors führt, obwohl dieser sehr vertraut erscheint. In einer Reihe von vier Experimenten wurden Kategorienamen als Hinweisreize verwendet, um die Wahrscheinlichkeit des *recall-to-reject* Prozesses zu erhöhen und damit falsche Erinnerungen zu reduzieren. Bei der Verwendung semantisch assoziierter Distraktoren in einem Item-Wiedererkennungstest (Experiment 1 und 2) wurde der Effekt falschen Wiedererkennens durch die Hinweisreize eliminiert. Allerdings konnte das absolute Niveau an Rekognitionsfehlern aufgrund einer gleichzeitigen Erhöhung der Vertrautheit und einer Tendenz der Probanden, konsistent zu antworten, nicht reduziert werden. Diese unerwünschten Nebeneffekte konnten durch die Verwendung phonologisch assoziierter Distraktoren erfolgreich ausgeschaltet werden: In Experiment 3 konnte die Unterdrückung des Effekts des falschen Wiedererkennens durch die Hinweisreize repliziert und die falsche Alarmrate sogar reduziert werden. Darüber hinaus zeigen die Ergebnisse von Experiment 1 und 3, dass der *recall-to-reject* Prozess auch spontan und ohne explizite Instruktion angewendet wird. Die Ergebnisse von Experiment 4 verdeutlichen, dass die Präsentation assoziierter Stimuli in der Lernphase eine effektive Anwendung dieses Prozesses verhindert. Die Ergebnisse der vorliegenden Arbeit liefern eindeutige Evidenz für den *recall-to-reject* Prozess und zeigen das Potential dieses Prozesses, falsche Erinnerungen zu reduzieren. Damit können sie als Evidenz für Zwei-Prozess-Theorien der Rekognition interpretiert werden.

Introduction

Memory researchers have long been interested in false memories. False memories are defined as memories for events that did not occur (Roediger & McDermott, 1995) or that occurred differently from the way in which they are remembered (Lindsay & Johnson, 2000). The great interest in false memories is understandable because studying false memories provides insights into general processes underlying memory (Reyna & Lloyd, 1997) and has important implications for real-life problems, like the accuracy of eyewitness testimony. In fact, real-life problems like the recovered memory debate of childhood sexual abuse (e.g., Memon & Young, 1997) dramatically intensified the study of false memories, which still enjoys great popularity. During the past decade there has been an increasing interest in memory editing mechanisms that can prevent false memories. However, evidence regarding memory editing mechanisms is rather confusing due to the enormous amount of experimental paradigms that have been employed for studying false memories. Additionally, numerous methodologies have been used to investigate processes underlying memory and discrepancies in terms and definitions exist. Not until very recently a first attempt has been made to link different phenomena and proposed memory editing mechanisms (Lampinen & Odegard, 2006).

The present work will focus on recall-to-reject, one of the central memory editing mechanisms in recognition memory. Recall-to-reject occurs when recall of a studied item is used to correctly reject an associated distractor despite its enhanced familiarity. For example, the word *mouse* might seem familiar if the word *rat* was presented in the study list, but if participants consciously recall that it was the word *rat* that was in the study list, they should reject the word *mouse* despite its high familiarity. Recall-to-reject thus reduces the typically observed heightened false alarm rate for associated distractors. Apart from its practical relevance to prevent false memories, the investigation of the recall-to-reject process is of great theoretical importance. Evidence of recall-to-reject is usually interpreted in favour of dual-process theories of recognition memory, which assume that recognition decisions are based on familiarity as well as on recollection processes (for an overview, see Yonelinas, 2002). However, evidence for recall-to-reject is far from clear-cut and most attempts to increase recall-to-reject have failed.

The theoretical part of the present work will begin with a short overview of recognition tasks that have been used to study false recognition and subsequent memory editing mechanisms thought to reduce false recognition. Second, a short overview of theories of recognition memory is provided with respect to their ability to account for false recognition errors and their compatibility with a recall-to-reject process. After a classification of memory editing mechanisms and the differentiation between recall-to-reject and other recall processes proposed to underlie recognition, the methods used to investigate recall-to-reject are outlined. Finally, evidence for the recall-to-reject process will be reviewed and moderating factors of recall-to-reject will be discussed. In the empirical part of the present work, the ability of category cueing to increase recall-to-reject and subsequently reduce false recognition will be investigated. Furthermore, the effect of explicit instructions and the importance of mutual exclusivity of stimuli for the effectiveness of recall-to-reject will be investigated.

1 False Recognition Tasks

It is a well-known empirical fact in recognition memory research that the false alarm rate for distractor items that are associated with studied items is systematically heightened relative to that of distractor items that are not associated with any studied item. This so-called false recognition effect can be observed for different kinds of associations and different recognition tasks. Generally, items that have been presented during study are termed targets, whereas items that have not been presented during study and therefore are new are termed distractors. Distractor items that are not associated with any studied item are often called standard distractors while distractor items that are associated with studied items are called associated distractors. The classification that will be provided in the following is important for the comparison of false recognition studies and, as will be explained in detail later, in particular for the study of recall-to-reject.

1.1 Item Recognition

The effect of association on recognition memory was first investigated with a continuous recognition task. Participants were presented with a list of words and had to indicate for each word, whether it had been presented earlier in the list or not. The false alarm rate was increased compared to standard distractors when items were semantically similar (Anisfeld & Knapp, 1968; Underwood, 1965) or phonologically similar (Anisfeld, 1969) to a previously presented item. A false recognition effect using the continuous recognition method has been found for multiple types of semantic similarity, for example for synonyms or antonyms of a studied word (e.g., always-forever; day-night), for common associative responses (e.g., bread-butter) or for category exemplars (e.g., birch-tree).

In contrast to continuous recognition, old/new recognition tests consist of a separate study and test list. During test, participants have to decide for each test item whether it has been presented in the study list or not. Beyond the semantic and phonological similarity effects already mentioned, a false recognition effect in old/new recognition has been demonstrated for plurality-changed nouns like *frog-frogs* (e.g., Hintzman & Curran, 1994), orthographic similarity of pseudo words such as *PRUMIR-PRUMAD* (e.g., Rotello & Heit, 1999), narrative statements (e.g., Brainerd, Reyna, & Estrada, 2006), and pictures (e.g., Curran & Cleary, 2003).

In forced-choice recognition tests, several items are presented at the same time and participants have to choose the item that has been presented during study. In the most simple form of forced-choice recognition, named two-alternative forced-choice test, only two items are presented during test (e.g., Tulving, 1981). To study the effect of association, a studied item is either paired with a standard distractor or an associated distractor and performance is compared between these two types of trials.

One of the most popular paradigms to study false memories was developed by Roediger and McDermott (1995) based on research by Deese (1959). In the Deese-Roediger-McDermott (DRM) paradigm, participants study lists of approximately twelve words (e.g., tired, dream, bed etc.) that are all semantically related to a nonpresented critical word (e.g., sleep), usually termed critical lure. Participants usually show high levels of false recognition as well as false recall for the nonpresented critical word. The false recognition rate can even reach the hit rate of studied words and the false recall rate approximates the recall rate of words presented in the middle of the study list (Roediger & McDermott, 1995). Apart from the high rate of false memories obtained with this paradigm, the remarkable persistence and participants' high confidence contribute to the great popularity of the DRM paradigm. If remember/know judgments (Tulving, 1985) are required, where participants have to indicate whether the acceptance of an item is based on the recollection of details of its presentation (remember-response) or on an unspecific feeling of familiarity (know-response), the percentage of remember-responses to nonpresented critical lures exceeds 50 percent (Roediger & McDermott, 1995). Participants even report details of presentation with high confidence (Payne, Elie, Blackwell, & Neuschatz, 1996). False memories in the DRM paradigm persist across long retention intervals (Seamon et al., 2002) and even explicit warnings cannot prevent the occurrence of false memories (Gallo, Roberts, & Seamon, 1997). The DRM paradigm has been successfully applied to different types of association and stimuli, such as category exemplars presented as words or as pictures (Seamon, Luo, Schlegel, Greene, & Goldenberg, 2000) or phonologically related words (Sommers & Lewis, 1999).

A standard old/new recognition test with non-associated study items and the DRM paradigm with up to 15 associated items are presented during study represent two extremes along a continuum. The majority of studies can be located in between these two extremes. The effect of the number of associated items presented during study has been investigated systematically. It has been shown that the false recognition effect increases with the number of associated items presented in the study phase (Dewhurst, 2001; Dewhurst & Anderson, 1999) and that presenting multiple associated items can lead to detailed illusory memories (Dewhurst & Farrand, 2004).

1.2 Associative Recognition

Recognition tasks that require the use of specific associative or contextual information have to be distinguished from standard item recognition tasks. Whereas item recognition requires discrimination of studied words from nonstudied words, associative recognition requires participants to discriminate studied word pairs from recombinations of words from different pairs. In a typical associative recognition task (e.g., Rotello & Heit, 2000), participants study pairs of words (e.g., A-B, C-D) and are tested on intact word pairs (A-B or C-D), rearranged word pairs (e.g., A-D), and completely new word pairs (e.g., X-Y). The false recognition effect in associative recognition is measured by the difference in false alarms between rearranged and new word pairs.

The memory conjunction paradigm is another type of associative recognition task that has become increasingly popular especially for the study of recall-to-reject (T. C. Jones, 2006; T. C. Jones & Atchley, 2006; Lampinen, Odegard, & Neuschatz, 2004; Odegard, Lampinen, & Toglia, 2005). In the memory conjunction paradigm, participants study compound words, which are termed parent words (such as *blackmail*, *jailbird, crossbow*). During test participants make recognition decisions for parent words (*blackmail*), conjunction items (*blackbird*), where both parts of the word are old but recombined, feature items (*rainbow*), where one part of the word is new and the other part is old, and completely new items. The pattern of false alarms typically found reflects the amount of similarity to studied items: The false alarm rate for conjunction items is higher than that for feature items and the latter is higher than that for new items (e.g., T. C. Jones & Atchley, 2006).

The distinction between item and associative recognition is important with respect to the investigation of recall-to-reject for several reasons. For example, the probability for recall-to-reject might per se be higher in associative compared to item recognition tasks. As an associated distractor in an associative recognition task is composed of a mixture of features from studied items, the familiarity of both targets and associated distractors should be very high. Therefore, recall of studied items should be particularly important for the discrimination between both item types and consequently the probability for recall-to-reject should be increased.

1.3 Discrimination Tasks

In discrimination tasks, participants are required to discriminate studied items with respect to the study context. Thus, associated distractors are items that have been studied but in the wrong context. In list discrimination tasks, participants study different lists of items and during test are required to respond old only to items from a specific list. For example, participants study items on list A and list B and are later asked to respond old only to items from list A and to respond new to unstudied items as well as studied items from list B. Other discrimination tasks require participants to discriminate between items presented in different modalities (e.g., Light, LaVoie, Valencia Laver, Albertson Owens, & Mead, 1992) or between items presented in a different format (e.g., Gallo, Cotel, Moore, & Schacter, 2007).

Discrimination tasks are source-monitoring tasks (Johnson, Hashtroudi, & Lindsay, 1993), in that participants have to discriminate items from different sources. Besides their use in source memory research, the discrimination task has been employed to separate automatic and consciously controlled processes underlying recognition memory. In the process dissociation procedure (Jacoby, 1991; Jacoby, Toth, & Yonelinas, 1993) the discrimination task described so far is called *exclusion* condition because studied items from one study context have to be excluded in the recognition test. The exclusion condition is contrasted with an *inclusion* condition in which participants are required to respond old to all studied items–irrespective of the study context–and to respond new only to unstudied items. It is assumed that an item will be rejected in the exclusion condition if participants recollect that the item is from the source to be excluded but the same item will be accepted in the inclusion condition. Thus, the process dissociation procedure implies that recall can be used for the rejection of items and discrimination tasks have recently been used to investigate recall-to-reject (Gallo, Bell, Beier, & Schacter, 2006; Gallo et al., 2007).

However, the comparison of results from discrimination tasks with the results of item or associative recognition tasks is problematic for two reasons. Firstly, it is questionable to what extent recall-to-reject that refers to the recall and rejection of a studied item is comparable to recall-to-reject that refers to the rejection of an unstudied item. In fact, as the former refers to the conscious recall of a studied item that has to be rated itself as being old or new, this process seems to be more equivalent to the recall-to-accept process (where recall of a studied item results in a hit) with the only difference that the item is not accepted but rejected as required by instructions. Secondly, the probability for recall-to-reject should be relatively high in a discrimination task because the recall as well as the rejection refer to the same studied item. By contrast, the probability of recall-to-reject should be lower in an item recognition task where rejection refers to an unstudied item based on the recall of the corresponding studied item.

To summarise, the classification provided (see Smith, Tindell, Pierce, Gilliland, & Gerkens, 2001 for a similar distinction) takes the differences between associated distractors in the different kinds of tasks into account. An associated distractor in an item recognition task is similar to a studied item but has not been studied itself. By contrast, an associated distractor in an associative recognition task is composed of a mixture of features from studied items. Finally, an associated distractor in a discrimination task has been studied but in a different context. These differences have important implications for the study of recall-to-reject. For example, comparisons between the different types of tasks are problematic since they are likely to differ with respect to the probability of the recall-to-reject process.

2 Theories of Recognition

There are many different theories about the processes underlying recognition decisions and most can explain the false recognition effect without difficulty. However, some theories are closely tied to the specific paradigms and materials used. A detailed presentation of all existing theories is beyond the scope of the present work. Instead, the classification of theories into single versus dual-process theories will be presented. This distinction is important for the present work as evidence of recall-to-reject is interpreted in favour of dual-process theories of recognition memory.

2.1 Single-process Theories

All theories of recognition memory which assume that recognition decisions are based on a single memory process, usually termed familiarity or memory strength, can be classified as single-process theories. As outlined in signal detection theory, it is assumed that test items are accepted as old if memory strength rises above a subjective response criterion (Snodgrass & Corwin, 1988). Different theories agree that the increase of the false alarm rate for associated distractors is due to an increase in familiarity. However, they differ in their assumptions about the processes underlying this increase of familiarity. For example, associated words may already become familiar during encoding because they are implicitly generated as a response to the presented word (Lampinen, Meier, Arnal, & Leding, 2005; Roediger & McDermott, 1995; Underwood, 1965) or because the activation of a presented word spreads throughout a semantic network (Anderson, Bothell, Lebiere, & Matessa, 1998; Collins & Loftus, 1975). Theories which assume that the familiarity increase of associated distractors is due to processes during encoding imply that memory representations for nonpresented items already exist. Thus, they can easily explain the false recognition effect for stimuli like words or common objects, but have difficulty to explain the effect for abstract material such as pseudo words (Rotello & Heit, 1999), dot patterns (Posner & Keele, 1970) or chessboard like patterns (Brandt, 2001).

By contrast, theories that assume the familiarity increase for associated distractors to be due to retrieval processes during recognition can explain the false recognition effect for familiar as well as abstract stimuli. It is assumed that the familiarity of a test item results from a comparison of the test item with the contents of memory. The fundamental idea of the so-called global memory models (e.g., Gruppuso, Lindsay, & Kelley, 1997) is that a test item serves as retrieval cue and is matched with *all* memory representations. The resulting familiarity is assumed to depend on the overall strength of all matches. This assumption implies the total similarity principle that predicts recognition decisions to be monotonically related to the total similarity of the test stimulus to memory for all stimuli. As an associated distractor is similar to at least one studied item whereas a standard distractor is not, the false recognition effect can be easily explained. Global memory models can be classified as single-process models because they assume that recognition memory judgments are based on a single mechanism (Gruppuso et al., 1997). However, models differ in their assumptions about the underlying encoding and retrieval processes. For example, SAM (Search of Associative Memory, Gillund & Shiffrin, 1984) and MINERVA 2 (Hintzman, 1988) are multiple trace models because they assume that each presentation of a stimulus results in a separate memory trace. By contrast, TODAM (Theory of Distributed Associative Memory, Murdock, 1982) is a composite vector model. Memory traces are assumed to be overlaid and combined into a single composite vector.

The success of global memory models can be explained by the fact that many phenomena can be explained with only few and simple assumptions. Furthermore, global memory models are mathematical models that formally describe the postulated encoding and retrieval processes. This allows for deriving exact and testable predictions. For example, the total similarity principle (see e.g., C. M. Jones & Heit, 1993) predicts that recognition as well as frequency judgments for a nonpresented item should monotonically increase with the number of presentations of similar items. This should be true even for the repeated presentation of a single similar item. However, both predictions could not be confirmed (for details see Chapter 5.1). Alternatively, the results could be explained by the contribution of a second recall-like process and therefore have been interpreted in favour of dual-process theories of recognition memory.

2.2 Dual-process Theories

Dual-process theories of recognition memory assume that recognition decisions are based on familiarity as well as on recollection processes (for an overview, see Yonelinas, 2002). Both memory processes are assumed to be functionally independent and qualitatively different. Familiarity is assumed to be a fast, automatic process that produces a feeling that the item was presented in the absence of the ability to consciously recollect it. Recollection refers to a slower and controlled retrieval process that results in the conscious recollection of details of its presentation. For targets, familiarity and recollection work in concert. Therefore, both processes can lead to a hit. However, for associated distractors, both processes oppose each other. Familiarity increases false alarms but recollection can be used to avoid the errors. Therefore, evidence for a recall-to-reject process can be interpreted in favour of dual-process theories.

A variety of dual-process theories exists, some of which differ in their conception of the interplay between the two processes assumed to contribute to recognition decisions. For example, a two-high threshold model (Atkinson & Juola, 1974) assumes that participants first evaluate the global familiarity of an item, with the item being accepted as old if familiarity exceeds some high subjective criterion or being rejected as new if familiarity falls below some low criterion[1]. Recollection is assumed to be a slow back-up process that searches memory only at intermediate levels of familiarity, resulting in an acceptance if producing a match between the test item and a memory trace and rejection if producing a mismatch. In contrast, Mandler (1980) postulated that the familiarity and recollection processes occur conjointly and that recognition involves the additive effects of familiarity and recollection.

A theory of recognition memory that has become increasingly popular, especially with respect to the investigation of the recall-to-reject process, is the fuzzy trace theory (Brainerd & Reyna, 2002a; Brainerd, Reyna, & Kneer, 1995). Central to the fuzzy trace

---

[1] A differentiation between the process of detection of a new item as proposed by two-high threshold models and the recall-to-reject process is provided in Chapter 3.1.2.

theory is the assumption of two independent memory traces for studied items. The gist trace is thought to represent the memory for the meaning of an item, whereas the verbatim trace is thought to represent the memory for surface properties of the item, such as perceptual details of its presentation. The two retrieval processes assumed in fuzzy trace theory are roughly comparable to the retrieval processes proposed by two-process theories of recognition memory (e.g., Jacoby, 1991; Mandler, 1980; Yonelinas, 1994, 2002). Activation of the gist trace corresponds to familiarity and activation of the verbatim trace corresponds to recollection. In contrast to fuzzy trace theory however, two-process theories allow the possibility that these two processes might work on a common memory representation (e.g., Gillund & Shiffrin, 1984). A recognition judgment based on the retrieval of the gist trace is termed similarity response whereas a judgment based on the verbatim trace is termed identity response. False recognition of associated distractors is assumed to result from gist based similarity judgments. Furthermore, it is assumed that the retrieval of the verbatim trace of the corresponding target will result in the rejection of an associated distractor. This process corresponds to the recall-to-reject process proposed by other dual-process theories. The recall-to-reject process will be defined and distinguished from related processes in the following Chapter.

3 Differentiation of Recall-to-reject and Related Processes

The recall-to-reject process has to be differentiated from other memory editing processes that can reduce false memories as well as other recall processes thought to underlie recognition memory. A differentiation from related processes is provided in the following paragraphs.

3.1 Memory Editing Processes

During the past decade there has been an increasing interest in mechanisms that minimise the occurrence of false memories. All memory editing processes in recognition are characterised by the fact that they involve some sort of recall of information, they are active in the sense that the information recalled it is acted upon, and finally, that they can oppose feelings of familiarity and hence reduce false recognition. Important differences between mechanisms exist though. However, the current state of research is confusing because numerous terms and definitions have emerged for identical mechanisms and identical terms have sometimes been used for different mechanisms. In the following paragraphs a classification is proposed and the four central memory editing mechanisms, which are currently discussed, will be presented.

3.1.1 Recall-to-reject

As already mentioned, associated distractors are often falsely accepted in recognition tests due to their enhanced familiarity. Recall-to-reject occurs when recall of a studied item is used to correctly reject an associated distractor despite its high familiarity. The example in the Introduction was that the word *mouse* will seem familiar if the word *rat* was presented in the study list. If the studied word *rat* can be recalled as such, however, the associated distractor *mouse* should be rejected despite its high familiarity. This process has been referred to in the literature as *recall-to-reject* (e.g., Rotello, 2001; Rotello & Heit, 2000), *disqualifying recall-to-reject* (Gallo, 2004), or *recollection rejection* (e.g., Brainerd, Reyna, Wright, & Mojardin, 2003; Lampinen et al., 2004). These terms are often used synonymously, although it has been argued that recollection rejection describes the underlying process whereas recall-to-reject the phenomenon observed (Odegard & Lampinen, 2006). Furthermore, the term recollection rejection is commonly used when data are interpreted in the context of fuzzy trace theory (Odegard & Lampinen, 2005). Finally, the term disqualifying recall-to-reject emphasises that the process is based on the recall of information which disqualifies the critical event from having occurred (Gallo et al., 2006). In the following the term recall-to-reject will be used.

The definitions of recall-to-reject differ depending on the underlying recognition tasks. In standard item recognition, recall-to-reject is defined as the rejection of a non-presented item due to the conscious recollection of the presentation of its corresponding target (Lampinen & Odegard, 2006). However, definitions in associative recognition and discrimination tasks encompass stimuli composed of a mixture of features of studied items and stimuli that have actually been studied albeit in a different context. In a broader definition, recall-to-reject is defined as the rejection of a suggested event due to the recollection of mismatching information (Gallo et al., 2007; Rotello, 2001). The methods that have been used to study recall-to-reject will be outlined in Chapter 4 and existing evidence of the recall-to-reject process will be presented in Chapter 5.

3.1.2 Identification

If multiple associated items are presented during study, an associated distractor can be rejected if participants realise during study that the item was missing on the study list. For example, this could occur within the DRM paradigm when participants think about the critical lure during study and later recall having thought about the item during study (Lampinen et al., 2005). The decrease of false recognition within the DRM paradigm that has been achieved by repeating study and test trials (Schacter, Verfaellie, Anes, & Racine, 1998) and forewarning participants (Gallo et al., 1997) has been discussed to result from an increased identification and subsequent rejection of the critical lure.

Beside the term *identification of the critical lure* (Starns, Hicks, & Marsh, 2006), an alternative account conceptualises this process as a variant of the recall-to-reject process (Lampinen et al., 2005). However, although both memory editing processes involve recall of information to reject an associated distractor, the main difference between them is that recall-to-reject is based on recall of information about the corresponding target whereas identification is based on recall of information about the associated distractor itself. Furthermore, identification has only been observed if multiple associated items are presented during study whereas–as will be discussed in detail in Chapter 5.5.2–recall-to-reject is most likely if the study list does not contain any associated items.

Please note that the term identification or detection of an item as new has also been used with respect to new items that are not associated with any studied item. However, this process that will be termed *detection* in the following, has to be differentiated from the memory editing identification process described. For example, the two-high threshold model of Atkinson and Juola (1974) assumes that the initial assessment of the global familiarity of an item will either result in the detection of a not presented item as new if the familiarity falls below some low criterion or in the acceptance of a studied item as old if familiarity exceeds some high subjective criterion. The fact that the detection of a new item is based on low familiarity is the crucial difference to the identification or recall-to-reject processes. Memory models differ with respect to the incorporation of such a process (e.g., Batchelder & Riefer, 1990; Bayen, Murnane, & Erdfelder, 1996). Models that do not incorporate the detection of a new item, usually assume that participants' responses to standard distractors are solely based on guessing.

3.1.3 Distinctiveness Heuristic

Another memory editing mechanism that has received great interest is the distinctiveness heuristic. The distinctiveness heuristic is a metacognitive strategy based on the assumption that people evaluate their memories against a criterion based on how vivid they expect their memories to be. An item is rejected if information that is expected cannot be retrieved. Besides being termed as *distinctiveness heuristic* (Dodson & Hege, 2005; Dodson & Schacter, 2001; Schacter, Israel, & Racine, 1999), this process has been referred to as *metamemorial suppression* (Brainerd & Reyna, 2002b), or *diagnostic recall-to-reject* (Gallo, 2004; Gallo et al., 2006). The term metamemorial suppression emphasises that it is a metacognitive strategy based on knowledge about one's own mental functioning. The term diagnostic recall-to-reject emphasises that the failure to recall to-be-expected information is diagnostic of the non-occurrence of an event.

Schacter and colleagues (1999) defined the distinctiveness heuristic as a mode of responding based on participants' metamemorial awareness that true recognition of studied items should include recollection of distinctive details. The idea was that encoding distinctive perceptual information could reduce the high levels of false recognition for the critical lure in the DRM paradigm (Israel & Schacter, 1997; Schacter et al., 1999). To test this hypothesis, either a picture or a visual word presentation accompanied the auditory presentation of lists of semantic associates. Picture encoding was found to result in fewer false alarms to critical lures compared to word encoding, which is indicative of a greater discriminability between studied items and related lures. The findings of Schacter and colleagues were replicated (Gallo et al., 2006) and extended (Dodson & Schacter, 2001; Gallo et al., 2007; Lloyd, 2007).

The distinctiveness heuristic has been challenged by two alternative explanations for the reduced false alarm rate for critical lures in the DRM paradigm after pictorial encoding. Firstly, it has been argued that the picture/word effect could simply be caused by a more conservative response bias (Gallo, Weiss, & Schacter, 2004). A second alternative explanation is the impoverished relational encoding hypothesis (Hege & Dodson, 2004). It is assumed that distinctive information interferes with the encoding of relational information which is the primary basis of false memories in the DRM paradigm. Consequently, the availability and the likelihood of coming to mind at retrieval or recognition should be decreased for critical lures after the encoding of distinctive information. Overall, there is less evidence for the criterion shift hypothesis and the impoverished relational encoding hypothesis than for the distinctiveness heuristic explanation (Dodson & Hege, 2005; Gallo et al., 2004).

Like recall-to-reject, the distinctiveness heuristic also depends on recall insofar as the judgment of the memorability of items depends on the recollection of studied items. However, whereas recall-to-reject is based on the presence of specific recollections, distinctiveness heuristic is based on the absence of expected recollections. Another important difference is the level at which the processes operate. The recall-to-reject process operates at the item level, that is, the rejection of an associated distractor is specifically based on the recollection of the corresponding target. The distinctiveness heuristic seems to operate at a global level because it is driven by global expectations about the vividness of memories formed during study (Dodson & Schacter, 2001; Gallo, 2004; Schacter et al., 1999).

3.1.4 Don't Recall-to-reject

There is another line of research that has been linked to the distinctiveness heuristic and sometimes has been quoted as evidence that the distinctiveness heuristic can also operate at the item level. However, the author believes that this line of research refers to a very similar metacognitive strategy that nevertheless has to be distinguished from the classical distinctiveness heuristic as well as from recall-to-reject. This related memory editing process is termed don't recall-to-reject.

The term *don't recall-to-reject* (Rotello, Macmillan, & Van Tassel, 2000) refers to a metacognitive process that occurs when an item is rejected if it is judged as memorable but lacks conscious recollection (e.g., see Ghetti, 2003; Rotello, 1999, 2001; Strack & Bless, 1994). This process has also been referred to as *memorability-based strategy* (Ghetti, 2008; Ghetti & Castelli, 2006), *metacognitive strategy* (Rotello, 1999; Strack & Bless, 1994), or *idiosyncratic distinctiveness heuristic* (Lampinen et al., 2005). For example, if your own name would be presented in a recognition test but was not on the study list, you would probably reject it with high confidence despite its high familiarity because you would be sure that you would have remembered if it had been presented. Usually, due to the limited capacity of memory and forgetting, the lack of conscious recollection is no valid criterion that an item has not been presented. However, the non-occurrence of an event is likely to be inferred from a lack of conscious recollection if metacognitive knowledge implies that the event would have been remembered if it had occurred.

The don't recall-to-reject process is very similar to the distinctiveness heuristic in that the lack of recollecting specific information results in a correct rejection. However, the inference made within the distinctiveness heuristic is based on the comparison of memorial evidence for the test item with memorial evidence for all studied items (Gallo, 2004). In contrast, the inference made within the don't recall-to-reject process is based on the fact that the test item is distinctive or salient and therefore judged as memorable (Howe, 1998). The differentiation proposed further implies that the distinctiveness heuristic operates at the global level and don't recall-to-reject is the corresponding process operating at the item level. This classification can explain some inconsistent findings. Evidence that was presented for the operation of the distinctiveness heuristic at an item level (Lampinen et al., 2005) actually applies to don't recall-to-reject.[2]

To summarise, the classification of memory editing mechanisms provided presented four different processes. Recall-to-reject occurs when recall of a studied item is used to correctly reject an associated distractor. The recall-to-reject process has to be differentiated form three other memory editing processes. Identification occurs if participants realise that an associated distractor has not been presented during study. The metacognitive strategy called distinctiveness heuristic occurs when an item is rejected because memorial evidence is not as vivid as is expected based on the recall of studied items. Finally, don't recall-to-reject occurs if a salient item is rejected because it is judged as memorable but not recalled. To complicate matters, other processes that involve recall have been proposed to underlie recognition. To differentiate recall-to-reject from these processes, the following paragraphs will give a short outline.

[2] The evidence was based on self reports in the DRM paradigm (Lampinen et al., 2005). The example provided for coding (One participant rejected the word *view* because he was sure he would have remembered that word because he had a good *view* in his old apartment.) demonstrates that the rejection was based on the participant's expectation of an individual item being highly memorable because of some personal importance of that particular item.

3.2 Other Recall Processes in Recognition

Apart from other memory editing processes, recall-to-reject has to be distinguished from other recall processes that have been proposed to contribute to recognition. Whereas the memory editing processes all result in the correct rejection of a distractor, the three other recall processes, which will be outlined in the following paragraphs, can be assigned to the remaining possible outcomes of an old/new recognition test. For an overview, Table 1 shows the possible outcomes of an old/new recognition test and the recall processes that these outcomes can be based on. Of course, recognition decisions can be based on familiarity only and do not necessarily have to include recall processes. All processes listed in Table 1 include conscious recall of detailed information. However, whereas recall-to-reject and recall-to-accept lead to true memories, erroneous recollection rejection and phantom recollection lead to false memories.

Table 1

Recall processes that have been proposed to contribute to recognition decisions and the respective possible outcomes in an old/new recognition test.

		Test item	
		Target	Distractor
Answer of participant	Old	Recall-to-accept → hit	Phantom recollection → false alarm
	New	Erroneous recollection rejection → miss	Recall-to-reject → correct rejection

3.2.1 Recall-to-accept

As discussed, dual-process theories assume that recognition decisions are based on familiarity as well as on recollection processes. If a correct response to a studied item is based on recollection, this process is referred to as recall-to-accept (Rotello, 2001; Rotello & Heit, 2000). Both recall-to-accept and recall-to-reject are based on a vivid memory for a target. However, whereas recall-to-accept occurs when a target is accepted (i.e., hit), recall-to-reject occurs, when an associated distractor is rejected (i.e., correct rejection).

3.2.2 Phantom Recollection

Phantom recollection occurs if false memory is accompanied by the illusory recollection of vivid and highly detailed information. Whereas recall-to-reject can lead to high-confidence rejections, phantom recollection can lead to high-confidence acceptances (i.e., false alarms) because the false memories are subjectively compelling. Evidence of phantom recollection (Brainerd, Wright, Reyna, & Mojardin, 2001; Dewhurst & Farrand, 2004; Lampinen et al., 2004) is important insofar as it demonstrates that false memories might not always be solely based on high familiarity levels.

False memories for critical lures in the DRM paradigm often meet the definition of phantom recollections because participants are highly confident and even report details of presentation (Payne et al., 1996; Roediger & McDermott, 1995). It has been suggested that phantom recollection results from content borrowing (Lampinen et al., 2005). It has been assumed that the high familiarity of the critical lure within the DRM paradigm leads to a search for episodic memory for details that would corroborate the feeling of familiarity. Finally, the vivid false memory occurs if episodic details of studied items are incorrectly attributed to the critical lure. Alternatively, phantom recollection may occur if associated distractors are confused with the corresponding target due to a very high similarity (Lampinen, Watkins, & Odegard, 2006).

3.2.3 Erroneous Recollection Rejection

There is evidence that the recall-to-reject process sometimes is inappropriately used for targets (Brainerd et al., 2003). This process termed erroneous recollection rejection can thus be defined as an inappropriate editing mechanism leading to the rejection of a true memory (i.e., miss). Whereas Brainerd et al. (2003) assume that erroneous recollection rejection is based on recollection of a related target, Lampinen et al. (2006) assume that it is based on recollection of a related distractor (i.e., phantom recollection). However, both explanations are not mutually exclusive. They agree that erroneous recollection rejection results in the rejection of a target and should not occur very often since a target should be a better retrieval cue for itself than for both, a related target and a related distractor.

4 Methods to Detect Recall-to-reject

As cognitive processes are not directly observable, different methods have been designed to demonstrate the existence of the recall-to-reject process. However, the results of different methods often are inconsistent, even when used simultaneously. Although manipulations of recall-to-reject are predicted to have an effect on the false recognition effect, most studies failed to find evidence for recall-to-reject based on recognition data. Instead, evidence based on multinomial modelling and ROC-analysis is numerous. To allow for the evaluation of evidence provided from different approaches, the basic principles of the different methods are described in the following paragraphs.

4.1 Recognition Data

Early evidence for the use of recall-to-reject stems from recognition data. A finding of Tulving (1981), termed test-pair similarity effect, violates the inverse relation typically found between recognition accuracy and the similarity between studied and new test items. Tulving's participants studied pictures and were then given a two-alternative forced-choice test. As expected, the hit rate was highest if a target item was paired with a dissimilar distractor. However, when targets were paired with a distractor similar to a studied item, unexpectedly, the hit rate was higher if the distractor was similar to the target item as opposed to another previously studied item. This result has been replicated using words (Hintzman, 1988) and has been taken as indirect evidence of recall-to-reject. The simultaneous presentation of target and associated distractor seems to draw participants' attention to distinguishing features, which facilitates the detection of mismatch between the associated distractor and the memory trace retrieved.

The test-pair similarity effect illustrates that recall-to-reject counteracts the effect of familiarity. With respect to old/new recognition, recall-to-reject should decrease the probability of an old response to associated distractors. As the recall-to-reject process is based on conscious recollection, enhancing the probability of conscious recollection should reduce the false recognition effect. Therefore, reducing the false recognition effect by manipulations that enhance recall can be taken as qualitative evidence for recall-to-reject. Accordingly, manipulations that reduce the probability of conscious recollection should result in an increase of the false recognition effect.

Another line of evidence for recall-to-reject based on recognition data has studied the time course of recognition decisions using the response signal procedure. In the response signal paradigm (e.g., Heit, Brockdorff, & Lamberts, 2004; Hintzman & Curran, 1994; T. C. Jones & Jacoby, 2001; Rotello & Heit, 1999; Rotello & Heit, 2000) participants make recognition decisions within some short interval following a signal to respond, with the signal being presented at various lags following the presentation of the test item. An inverted U function is expected for false alarms to associated distractors based on the assumption that familiarity is a fast acting rather automatic process and recall a rather time consuming more strategic process (Yonelinas, 2002). As long as responses are solely based on familiarity, the false alarm rate should increase with longer response lags as familiarity evidence accumulates. However, once the recall process has begun to contribute to recognition judgments, the false alarm rate is predicted to decrease with longer response lags.

As recognition decisions can be based on multiple cognitive processes (such as familiarity, guessing, and potentially recall processes), it is obvious that recognition data cannot provide a pure measure of a specific memory process. Therefore, interpretations should not be based on the false alarm rates for associated distractors only. Recognition data should rather be evaluated in relative terms. For example, with respect to the response signal paradigm, it seems plausible that response bias changes over processing time. However, it seems reasonable to analyse the difference between the false alarms for standard distractors and associated distractors, that is, the false recognition effect, which should not be influenced by response bias. Furthermore, multiple discrimination measures, such as d' (Heit et al., 2004; T. C. Jones, 2006; Snodgrass & Corwin, 1988), d_L (Hintzman & Curran, 1994; Rotello & Heit, 1999, 2000) or *pseudo* d_L (Rotello, 2001) have been used[3]. Moreover, the experimental design is crucial for the interpretation of recognition data.

---

[3] The discrimination/sensitivity measure $d' = z(H)-z(FA)$ is based on normal distribution, $d_L = \ln [H(1-FA)/FA(1-H)]$ is based on logistic distribution, where H is the probability of a hit and FA the probability of

As increasing the recall-to-reject process is predicted to result in a reduction of false recognition, recognition data are important with respect to theoretical as well as practical considerations. Recognition data will be used in the present work as the dependent variable. However, in contrast to recognition data that can provide qualitative evidence for recall-to-reject, quantitative estimates of recall-to-reject can be obtained using a mathematical model termed conjoint recognition.

4.2 Conjoint Recognition

The conjoint recognition model (Brainerd, Reyna, & Mojardin, 1999) is a multinomial processing tree model based on the assumptions made in fuzzy trace theory. Generally, the aim of multinomial processing tree models is to disentangle and measure the probabilities of different cognitive processes underlying the observed categorical data (Batchelder & Riefer, 1999; Riefer & Batchelder, 1988).

Like the process dissociation procedure (Jacoby, 1991; Jacoby et al., 1993), conjoint recognition is an opposition procedure designed to provide quantitative estimates of the processes underlying recognition memory. In conjoint recognition, participants study items and later complete a recognition test containing the three item types, targets, standard distractors, and semantically associated distractors. Three groups of participants are tested with different instructions. *Target instructions* correspond to standard instructions, as only studied items are to be accepted as old. Under *related instructions* participants are to respond old only to semantically associated distractors and under *target + related instructions* participants are requested to accept both targets and associated distractors.

Figure 1 depicts a multinomial processing tree of the conjoint recognition model to illustrate the basic ideas. Each branch of the tree represents a specific hypothesised sequence of cognitive processes, resulting in a specific response category. Typically, an

---

a false alarm. Pseudo d_L is used to indicate the ability to discriminate between standard and associated distractors.

observed response category can result from different sequences of cognitive processes. For instance, under target instructions, a new response is given when the verbatim trace of the corresponding target is retrieved. However, a new response can also be based on guessing when neither the verbatim trace of the corresponding target nor the gist trace is retrieved. The probability of a response category is expressed as a function of the underlying parameters. For instance, the probability for a new response to associated distractors under target instructions can be expressed as P(new) = V + (1-V)(1-G)B. By transforming the resulting equations, parameter estimates can be calculated based on the observed response frequencies.

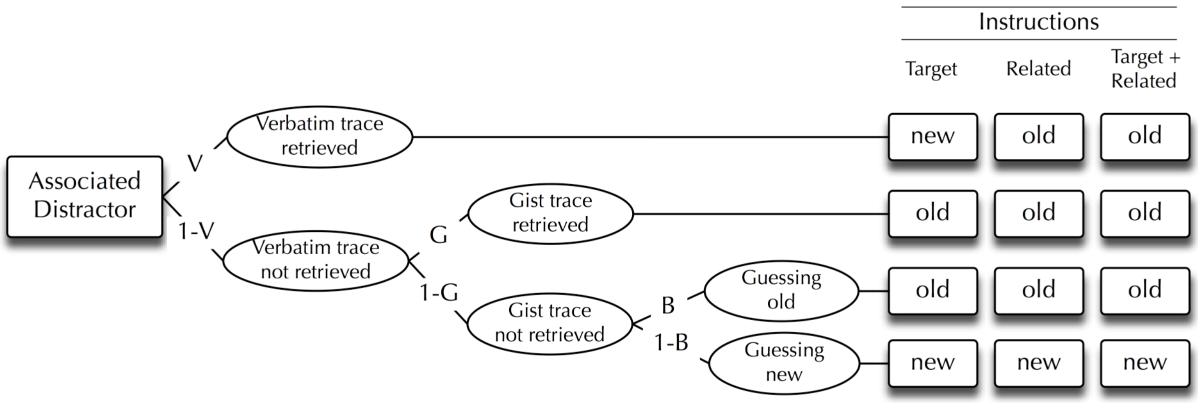

Figure 1 Multinomial processing tree diagram of the conjoint recognition model for associated distractors. The branches of the processing tree represent the (combination of) underlying cognitive processes (retrieval of the verbatim trace corresponds to recollection, retrieval of the gist trace corresponds to familiarity). The rectangles on the right represent the resulting response categories in the three different instruction conditions. The conjoint recognition model is composed of further tree diagrams for targets and standard distractors[4].

The conjoint recognition model includes a separate parameter to measure the recall-to-reject process (parameter V in Figure 1). The logic is to test participants' recognition memory under instructional conditions in which recall-to-reject should lead to different responses. As illustrated in Figure 1, when participants have to respond to associated distractors, the corresponding studied item will be recalled with the probability V. The recollection of the corresponding studied item results in a rejection under target instructions but in an acceptance under related as well as target + related instructions. This opposition is used to estimate the probability of recall-to-reject. The conjoint recognition model has been used to investigate recall-to-reject in a variety of studies (Brainerd & Reyna, 2002b; Brainerd et al., 2006; Brainerd et al., 2003; Brainerd, Stein, & Reyna, 1998; Odegard & Lampinen, 2005; Rotello, 2001; Rotello & Heit, 1999). The findings that have been obtained will be presented in Chapter 5.

A model termed simplified conjoint recognition has been developed recently to provide estimates of the same processes as conjoint recognition but with only a single group of participants (Stahl & Klauer, in press). Therefore, the simplified conjoint recognition model is much more efficient in terms of data collection than the original conjoint recognition model. In a nutshell, instead of an old/new recognition test, participants are asked to identify the type of test item and respond either with *target*, *related* or *new*.

4.3 ROC Analysis

Another method that has been used to estimate the occurrence of the recall-to-reject process is the extended analysis of receiver operating characteristic (ROC) curves (Lampinen et al., 2004; Rotello, 2001; Rotello et al., 2000). To obtain ROC curves, participants in an old/new recognition test additionally have to rate the confidence with which they made their recognition decisions. The traditional ROC curve (Yonelinas, 1994) is generated by plotting the cumulative hit rate for targets against the cumulative false alarm rate for standard distractors across confidence levels. Each point on an ROC plot gives the cumulative probability of assigning targets and standard distractors, respectively the same confidence value. The left-most point on the curve represents *sure old* judgments to targets (on the y-axis) and to new items (on the x-axis), the second point includes both *sure old* and *probably old* judgments, and so on. Thus, ROC curves are monotonically increasing. The shape of the ROC curve is predicted to

[4] The conjoint recognition model does not incorporate the process of detection of a new item (see Chapter 3.1.2). It is assumed that participants' responses to standard distractors are solely based on guessing.

depend on the underlying cognitive processes. If recognition is solely familiarity-based, the ROC curve is predicted to be curvilinear (and symmetric if normal distribution and equality of variances is assumed). In contrast, a linear ROC curve is predicted if recognition is assumed to be based on a single high-threshold recollection process. Dual-process views predict that the ROC curves will be asymmetric and increasingly linear as the relative importance of the recall component increases.

Estimates for recall-to-reject can be obtained by extending the traditional ROC analysis (Rotello, 2001; Rotello et al., 2000). The interpretation is based on the assumption that recall-to-reject results in a rejection of an associated distractor at the highest confidence level. As illustrated in the left part of Figure 2, the hit rate for targets is plotted against the false alarm rate for standard distractors (traditional old-new ROC curve) and, additionally, against the false alarm rate for associated distractors (old-similar ROC curve). The difference between the old-new and old-similar ROC curve with respect to the upper x-intercept is thought to reflect the amount of recall-to-reject. The logic is that if theoretically, the response criterion is maximally liberal so that all targets are accepted as old, *sure new* rejections of associated distractors have to be due to recall-to-reject. Therefore, the more *sure new* judgments are given for associated distractors based on recall-to-reject, the earlier will the old-similar ROC curve hit the upper x-axis, given that a *sure new* judgment is not made for targets (which could occur based on erroneous recollection rejection).

Similarly, it is assumed that a recall-to-accept process results in an acceptance of a target at the highest confidence level. Therefore, a y-intercept of both the old-new and old-similar ROC curves greater than zero is thought to reflect the amount of recall-to-accept (see Figure 2). The logic is that if theoretically, the response criterion is maximally conservative so that no judgments are based on familiarity (which is assumed to be driven to zero when the false alarm rate is zero), *sure old* judgments for targets have to be due to recall-to-accept.

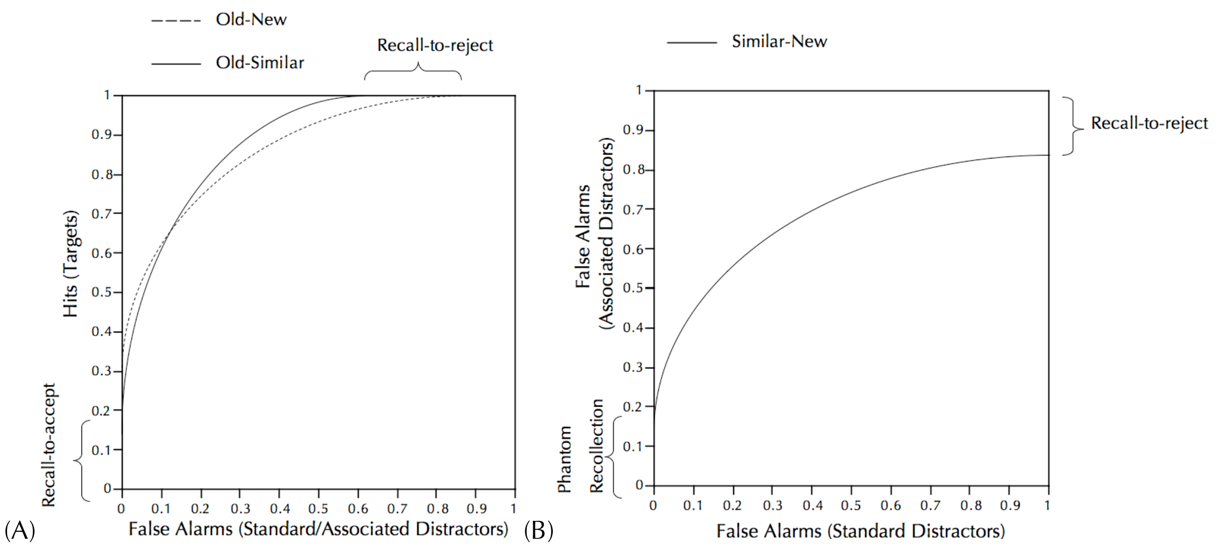

Figure 2 Hypothetical ROC curves for the extended ROC analysis proposed by (A) Rotello and colleagues (2001; 2000) and (B) Lampinen and colleagues (2004). For details see text.

Lampinen and colleagues (2004) have proposed another variant of ROC analysis that can provide estimates for recall-to-reject as well as phantom recollection. As illustrated in the right part of Figure 2, the false alarm rate for associated distractors is plotted against the false alarm rate for standard distractors, resulting in a similar-new ROC curve. The y-intercept of the similar-new ROC curve is thought to represent phantom recollection. The logic is the same as for recall-to-accept in old-new or old-similar ROC curves. The y-intercept represents the point at which no standard distractors are being accepted, and, hence familiarity has been driven to zero. Thus, any related lures that are accepted at that point should reflect phantom recollection. Recall-to-reject is thought to occur if the similar-new ROC curve intercepts the right y-axis below the point (1,1). The logic is that even when participants are being so liberal as to accept all standard distractors, the remaining rejections of associated distractors have to be due to recall-to-reject.

Furthermore, the phantom ROC model combines ROC analysis with the conjoint recognition model (Lampinen et al., 2006). Besides parameters for familiarity, recall-to-accept and recall-to-reject, the phantom ROC model includes parameters for phantom recollection and erroneous recollection (for a graphical illustration see Figure 3).

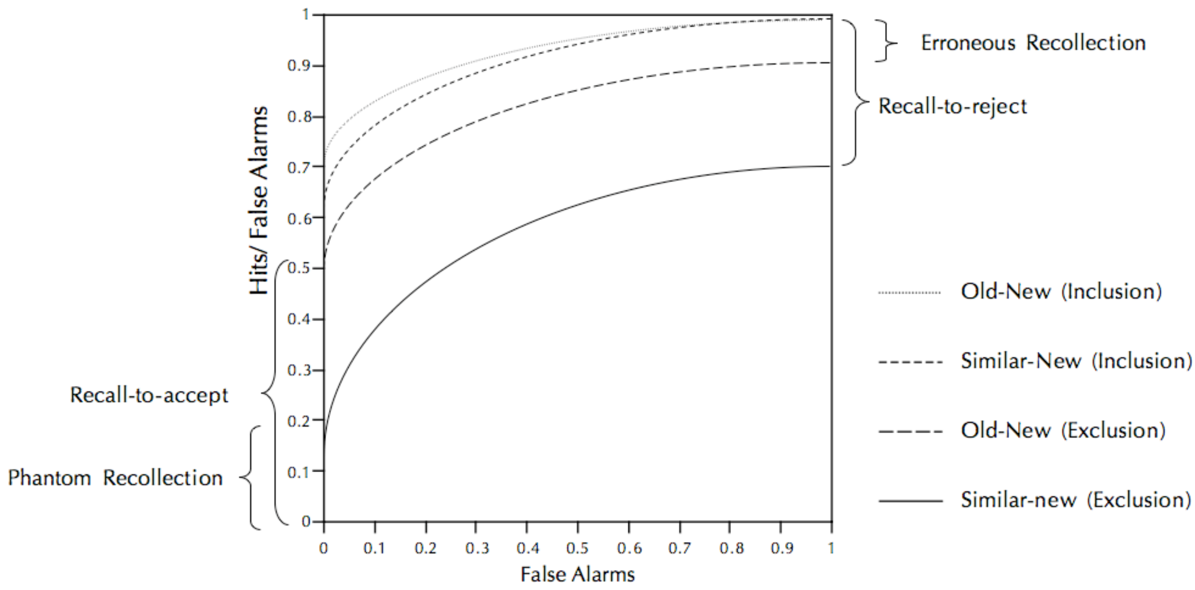

Figure 3 Hypothetical ROC curves predicted by the phantom ROC model (Lampinen et al., 2006) and differences that provide estimates for recall processes in recognition. For details see text.

In the phantom ROC model, participants are first requested to respond under inclusion instructions (i.e., accept targets and associated distractors) followed by a confidence rating. Secondly, participants are requested to respond under standard exclusion instructions (i.e., accept only targets), again followed by a confidence rating. Based on the logic that the recall of a target will result in a rejection of an associated distractor under exclusion instructions but in an acceptance under inclusion instructions, the differences of similar-new ROC curves under inclusion versus exclusion instructions provides a measure of recall-to-reject additional to the parameter estimate. The difference thought to reflect recall-to-reject corresponds to the right y-intercept as the similar-new ROC curve under inclusion instructions is expected to intercept the y-axis at the point (1,1). Identical to the ROC variant described before (Lampinen et al., 2004), the y-intercept of the similar-new ROC curve provides an estimate for the occurrence of phantom recollection. Estimates for recall-to-accept and erroneous recollection rejection are obtained from traditional old-new ROC curves. Identical to the ROC analysis of Rotello and colleagues (2001; 2000), the y-intercept of the old-new ROC curve provides an estimate for the occurrence of recall-to-accept. Similar to recall-to-reject, the differences of old-new ROC curves under inclusion versus exclusion instructions are thought to reflect erroneous recollection rejection. However, as the old-new ROC curve under inclusion instructions is expected to intercept the y-axis at the point (1,1), the difference corresponds to the right y-intercept.

4.4 Self Reports

It has also been proposed to use self reports to estimate the occurrence of recall-to-reject. For example, self reports can be obtained by asking participants to provide an explanation for each of their responses (Odegard et al., 2005) or to think out loud during the recognition test (Lampinen et al., 2005). Like confidence judgments which are part of ROC analysis, self reports are a type of introspective data. However, as introspection is a subjective and rather unreliable measurement tool, self report data have to be interpreted with caution. For example, it seems likely that participants report the use of recall-to-reject more often when they were given explicit recall-to-reject instructions. Usually, self reports are used in combination with other methods (Gallo, 2004; Lampinen et al., 2004; Odegard & Lampinen, 2005; Odegard et al., 2005). However, self report data are not always in line with other methods used to investigate recall-to-reject. For instance, in a study using the memory conjunction paradigm evidence for recall-to-reject was found only in self reports but not in old/new recognition data (Odegard et al., 2005).

5 Evidence for Recall-to-reject

Currently, there are diverse lines of evidence for the recall-to-reject process. Existing evidence differs with respect to the methods used to detect recall-to-reject and the manipulations that have been used to influence the probability of recall-to-reject. The rationale of all manipulations is that the probability of recall-to-reject is tied to the probability of conscious recollection. Enhancing the probability of conscious recollection should increase the probability of recall-to-reject and consequently reduce the false recognition effect. Therefore, reducing the false recognition effect by manipulations that enhance recall can be taken as qualitative evidence for recall-to-reject. Accordingly, manipulations that reduce recall should increase false recognition, although this direction of manipulation is less important with respect to practical considerations.

Basically, conscious recollection of a target corresponding to an associated distractor should depend on the availability as well as the accessibility of the memory representation (Tulving & Pearlstone, 1966). Availability refers to the existence of a memory representation of an experienced event and can be varied by manipulations during the study phase. Accessibility refers to the ability of a person to gain access to a stored memory trace and can be varied by test phase manipulations. This distinction between study and test phase manipulations will be used in the following to provide a better overview of manipulations that have been implemented. However, indirect evidence based on judgments of frequency or the response signal procedure will be presented beforehand. Both lines of evidence could also be assigned to the study and test phase distinction, but the differences in methods suggested a separate presentation.

5.1 Judgements of Frequency

Besides the test-pair similarity effect, further indirect evidence for recall-to-reject stems from studies in which participants have to estimate the frequency of items. Old/new recognition can be taken as a special case of judgments of frequency. In old/new recognition, a new response corresponds to a frequency judgment of zero and an old response corresponds to a frequency judgment greater than zero.

Frequency judgments have been used to test the total similarity principle that is implied in global memory models. It is assumed, that recognition and frequency judgments are monotonically related to the total similarity of the test stimulus to memory for all stimuli. Therefore, the total similarity principle predicts that even frequency judgments for a nonpresented item should monotonically increase with the number of presentations of similar items and even the repeated presentation of a single similar item.

However, this prediction could not be confirmed by C. M. Jones and Heit (1993) and only partially by Hintzman, Curran, and Oppy (1992). In the study of Hintzman and colleagues, about half of the frequency judgments were zero judgments and the predicted monotonic increase with the number of presentations of the similar word could only be observed if zero judgments were removed. Associated distractors in this study were mirror reversals of photographs and line drawings or plurality-changed nouns and therefore highly similar to the presented item. Hintzman and colleagues hypothesised that zero judgments were made despite the high familiarity if participants realised the mismatch between the test item and the corresponding presented item.

Contrary to predictions, no effect of the number of presentations on frequency and recognition judgments for an associated distractor was found in the study of C. M. Jones and Heit (1993). Semantically similar words such as *creek-stream* were used as stimuli in this study. Additionally, the frequency judgments for presented items can be interpreted with respect to recall-to-reject. As predicted, frequency judgments for presented items increased with the presentation of a similar word during study. However, unexpectedly, if the similar item presented during study was repeated more than six or twelve times (Experiment 3 and 4, respectively) the mean frequency estimate decreased again. The high number of presentations of the similar presented word could have triggered recall, therefore leaving the frequency estimate of the similar test item unaffected.

5.2 Response Signal Procedure

Based on the assumption that familiarity is faster than recollection, the false recognition effect is predicted to increase for fast judgments due to the familiarity based process and to decrease later in processing due to the contribution of recall-to-reject (e.g., Rotello, 2001). However, evidence for the recall-to-reject process using the response signal procedure is rather weak.

Contrary to predictions, the delay of the response signal had no effect on false recognition in an item recognition paradigm (Rotello & Heit, 1999) and the memory conjunction paradigm (T. C. Jones, 2006; T. C. Jones & Jacoby, 2001). At first glance, the biphasic false alarm rate for associated distractors, which were plurality-changed nouns, seemed to provide evidence for the response signal prediction (Hintzman & Curran, 1994). However, the decrease at longer response lags seemed to be driven by a more conservative response bias, as a reanalysis of the data revealed that the discrimination measure d_L did not change (Rotello & Heit, 1999). Similarly, the false alarm rate to critical lures in the DRM paradigm was found to decrease over longer response lags (Heit et al., 2004), but false alarms to standard distractors also decreased. The size of the false recognition effect was not analysed, but seemed quite constant at a descriptive level.

Mixed evidence for recall-to-reject has been provided using the response signal procedure in a discrimination task (T. C. Jones, 2006). When participants had to discriminate between read and self generated words and the response signal delay varied between 250 ms and 2250 ms, discrimination did not improve until the longest delay. However, within a single study, the predicted peaked pattern was observed analysing the false recognition effect and calculating d_L in an associative recognition task, but not in an item recognition task (Rotello & Heit, 2000). Further evidence for recall-to-reject in associative recognition has been reported by Rotello and colleagues (2000). Analysing ROC curves provided evidence for recall-to-reject after a long response delay (2500 ms) but not after a short response delay (450 ms).

5.3 Study Phase Manipulations

Although many established methods to influence recall by manipulations during the study phase exist, their success in affecting the recall-to-reject process is not as robust as might be expected.

5.3.1 Level of Attention

Dividing attention at study should decrease participants' ability to store memory traces for studied items and consequently to use recall of studied items to reject associated distractors. Dividing attention during study has been found to reduce the recall-to-reject parameter of phantom ROC (Lampinen et al., 2006) and conjoint recognition (Odegard & Lampinen, 2005). However, in the latter study, the level of attention did not have an effect on false recognition. Yet, the data have to be interpreted with caution, as a false recognition effect was neither observed in the full nor in the divided attention condition. Additionally, dividing attention had no effect on false recognition in the memory conjunction paradigm (T. C. Jones & Jacoby, 2001).

5.3.2 List Length

Increasing the length of study lists should decrease the quality of memory traces for studied items and consequently decrease rates of recall-to-reject. Although the hit rate for targets has been found to be higher after a short study list (40 words) compared to a long study list (80 words), the amount of false recognition in the memory conjunction paradigm was not influenced by the length of study lists (Odegard et al., 2005).

5.3.3 Repetition

Early studies of experimental memory research showed that study repetition increases recall performance (Ebbinghaus, 1964). Accordingly, repeating items during study increases the hit rate for targets in recognition memory (Stahl & Klauer, in press; Tussing & Greene, 1999). Consequently, repeating items during study should increase recall-to-reject and therefore reduce false recognition.

A discrepancy between recognition data and model parameters has been found in some studies investigating the effect of repetition of items during study. In the memory conjunction paradigm, presenting targets three times during study compared to only once increased the recall-to-reject parameter of ROC analysis and the use of recall-to-reject measured by self reports but did not reduce the false recognition effect (Lampinen et al., 2004). Similarly, a discrepancy between the recall-to-reject parameter of the conjoint recognition model and recognition data has been found for the recognition of narrative statements (Brainerd et al., 2006). In a study that used semantically related words as associated distractors, presenting targets twice has also been shown to increase the recall-to-reject parameter of the conjoint recognition model, but unfortunately recognition data were not provided in this study (Brainerd et al., 1999). In contrast to the conjoint recognition model and ROC analysis and contrary to predictions, the recall-to-reject parameter of the simplified conjoint recognition model was not affected by repeating targets twice during study (Stahl & Klauer, in press).

A null effect of repetition with respect to recognition data has been reported by T. C. Jones and Jacoby (2001) and Tussing and Greene (1999). These results are in line with the null effect of similar word frequency on the judgment of frequency for associated distractors (C. M. Jones & Heit, 1993) that has been discussed in Chapter 5.1.

5.4 Test Phase Manipulations

Manipulating the delay between study and test is presented along with other test phase manipulations, because the time of test is varied and it has been assumed that an increasing delay reduces the accessibility of memory representations.

5.4.1 Delay

Recall performance is usually found to decrease with increasing delay between study and test (e.g., Seamon et al., 2002). Consequently recall-to-reject should also decrease and lead to an increase in false recognition. The same prediction is made by fuzzy trace theory which assumes that, as time passes, memory for the surface details of experience becomes inaccessible more rapidly than memory for the gist of experience (Brainerd et al., 2003). The finding that the recall-to-reject parameter of the conjoint recognition model has been found to be above zero on immediate tests but decreases to zero after a one week delay is consistent with this prediction, although recognition data were not presented (Brainerd et al., 2003). Combining continuous recognition with the memory conjunction paradigm, T. C. Jones and Atchley (2002, 2006) manipulated the lag between the respective targets and the conjunction lure. Although a lower conjunction error rate was expected for relatively short lags, lag had little effect on conjunction errors unless it was reduced to zero (T. C. Jones & Atchley, 2002). The same initial increase of conjunction error rates from a lag of zero to one word and unexpectedly, a decrease from a lag of one word onwards was found in a very similar study (T. C. Jones & Atchley, 2006). Only the initial increase of conjunction errors is consistent with the hypothesis that false recognition should increase with increasing delay. However, a lack of a further increase might be due to the fact that participants' time to respond was restricted to 1500 ms. Furthermore, within the DRM paradigm, the amount of self reported recall-to-reject has been found to be lower after a 48 hour delay between study and test (Lampinen et al., 2005).

5.4.2 Target Priming

Brainerd, Reyna and Kneer (1995) argued that the best retrieval cue is the target itself and therefore presented targets just prior to their associated distractors. This target priming technique turned out to be a simple and powerful method to reduce false recognition errors. The false alarm rate for associated distractors was reduced down to or even below the level of standard distractors, an effect that has been labelled false recognition reversal. Target priming also increased the recall-to-reject parameter of conjoint recognition (Brainerd et al., 1999) and simplified conjoint recognition (Stahl & Klauer, in press). In another study (Brainerd et al., 1998), the effect of target priming was limited to the recall-to-reject parameter of the conjoint recognition model, an effect on recognition data could not be observed. This data have to be interpreted with caution though, as no false recognition effect could be observed in the baseline condition.

However, the high effectiveness of the target priming technique seems to be partly due to a response strategy (Wallace, Malone, Swiergosz, & Amberg, 2000). If the study list does not contain any pairs of related items, it seems reasonably to accept no more than one word of a related word pair during test. Therefore, an associated distractor should be rejected if the preceding target has already been accepted. The rejection will only be due to recall-to-reject if the preceding target has been recalled. However, if the preceding target has been accepted based on high familiarity or guessing, the rejection of the following associated distractor will be purely strategic. Such a response strategy works in line with recall-to-reject. Therefore, the unexpected reversal of the false recognition effect (Brainerd et al., 1995) could be based on the contribution of such a response strategy. Similarly, evidence for recall-to-reject within the conjoint recognition model that has been found even though no false recognition effect could be observed in the baseline condition (Brainerd et al., 1998) could be due to the contribution of strategic processes.

5.4.3 Cueing

Another possibility to enhance recall during the test phase might be the presentation of category names as retrieval cues. When participants studied category exemplars as items, Tulving and Pearlstone (1966) could show that presenting the corresponding category names as retrieval cues in a recall test increased the number of items recalled. As the participants' task was to reproduce learned items, the better performance in cued compared to free recall must have been due to an improvement of conscious recollection. Hence, if presenting category names did improve the recall of the corresponding learned category exemplars, it might also be possible to use this effect in a recognition task. Presenting category names as retrieval cues should increase the probability of the recall-to-reject process and consequently reduce the false recognition effect.

The category cueing manipulation will be used in the present work. A first attempt to use category cues to reduce false recognition was already made by Gallo (2003, 2004). In a pilot experiment category names were either presented along with category exemplars during study and/or at test. However, although the presentation of category names did improve recall, no effect on recognition was observed. Several factors might be responsible for this lack of an effect. Firstly, the presentation mode of the cueing manipulation might have impaired the effectiveness of retrieval cues. As category names were merely presented along with category exemplars, participants may have ignored the category names since they were unnecessary for solving the task. Also, the required recognition judgment was much less demanding and time consuming than it would have been to additionally engage in recall processes concerning the category names. Secondly, the number of exemplars learned per category was five and thus presumably too high for effectively utilizing recall-to-reject. There is evidence that recall-to-reject requires mutual exclusivity of items or exhaustive recall of all associated items learned which becomes less probable as the number of associated items increases (Brainerd et al., 2003; Gallo, 2004). The importance of mutual exclusivity and exhaustive recall, as well as other moderating variables will be discussed in the following paragraphs.

5.5 Moderating Factors for Recall-to-reject

As presented in the previous Chapter, the evidence for recall-to-reject is mixed and identical manipulations have produced inconsistent results in different studies. However, some inconsistencies can be explained by moderating variables.

5.5.1 Instructions

In most experiments investigating recall-to-reject, participants are given instructions that explicitly define the relationship between study items and associated distractors and explain the nature of test list construction. For example, the *related* and the *target + related* instructional conditions of the conjoint recognition model require explicit instructions about the different types of distractors and therefore automatically call participants attention to associated distractors. Additionally, explicit instructions are often given about how the recall of studied items can be used to avoid false memories for associated distractors (Gallo et al., 2007; Lampinen et al., 2004; Odegard & Lampinen, 2005; Rotello, 2001). For example, in an item recognition task, participants were informed that the test list would include studied items, new words and words that differed in plurality from the studied items. They were additionally told that only the singular or plural form of a word was studied, so that if they could remember studying the word *frog*, for example, they could be sure that *frogs* had not been studied (Rotello, 2001). Explaining the experimental paradigm and explicitly instructing participants to use the recall-to-reject process has been shown to reduce the false recognition effect (Gallo, 2004; Lampinen et al., 2004) and to increase the recall-to-reject parameter of ROC analysis (Lampinen et al., 2004; Rotello et al., 2000).

Only few studies exist where no explicit instructions were given and these studies failed to provide evidence for recall-to-reject (Rotello & Heit, 1999; Tussing & Greene, 1999). Thus, it is questionable how much previous findings were driven by providing participants with explicit instructions about the nature of the study and test environment. It has been argued that previous findings might be artefacts created by the methodologies used and the need for more research on the role of instructions has been emphasised (Lampinen et al., 2006; Odegard & Lampinen, 2006). However, indirect evidence for recall-to-reject exists (e.g., see Chapter 5.1). Thus, individuals seem to be capable of identifying the test structure themselves and of spontaneously using recall-to-reject. One goal of the present work was to find evidence for recall-to-reject without explaining the study and test list structure and without giving explicit recall-to-reject instructions.

It is still an unresolved question whether recall-to-reject is an automatic process that occurs outside of conscious awareness or a rather controlled and strategic process that requires conscious awareness. The finding that explicit instructions increase the use of recall-to-reject is in line with the assumption that the level of recall-to-reject increases with participants' awareness of the mechanism (Brainerd & Reyna, 2002b). However, it remains unresolved whether recall-to-reject is restricted to conscious awareness or can occur automatically based on a feeling of contrast between the presented and nonpresented items, as it is assumed by Brainerd and colleagues (2003). Evidence for the necessity of mutual exclusivity of stimuli for the use of recall-to-reject that will be presented in the following paragraph argues for the strategic character of the recall-to-reject process.

5.5.2 Mutual Exclusivity

The example for recall-to-reject used in the Introduction of this work was that the recall of the studied item *rat* could be used to correctly reject the associated distractor *mouse*. This conclusion can be made logically if the study list does not contain subsets of semantically related words. However, it has been discussed whether recall-to-reject can still be used if associated items are not mutually exclusive. Rotello and Heit (1999) directly tested the effect of mutual exclusivity. Some participants studied both items of pairs of similar pseudowords (e.g., *PRUMIR* and *PRUMAD*). In the mutual exclusive condition, participants always studied only one item of pairs of similar pseudowords. However, Rotello and Heit failed to find evidence for recall-to-reject in any condition. Thus, mutually exclusive stimuli are not sufficient for recall-to-reject, but the question remained if mutual exclusivity is a necessary condition for the use of recall-to-reject.

As recognition tasks differ with respect to mutual exclusivity, comparing the evidence for recall-to-reject from different types of tasks might shed light on this question. As semantically related words are presented in the DRM paradigm, evidence for recall-to-reject should not be found within the DRM paradigm if mutual exclusivity of stimuli is a necessary precondition. Evidence for recall-to-reject in the DRM paradigm has been reported (Lampinen et al., 2005), but evidence was only based on self reports in this study. Secondly, the finding that false recognition for critical lures in the DRM paradigm was reduced if critical lures were *tagged* was interpreted as evidence for recall-to-reject (Gallo et al., 2006). But as tagging critical lures meant that they were presented on a separate list and participants were instructed not to accept these critical lures, the task was turned into a discrimination task. Critical lures are no longer nonpresented items, but items presented on an exclusion list. Thus, although multiple associated items are presented in the study list, tagging critical lures makes them mutually exclusive in that they can logically be excluded if participants recall their presentation on the exclusion list. Thirdly, the effect of mutual exclusivity on recall-to-reject has been tested using a discrimination task, in which participants had to decide whether object names were additionally presented in red font or as pictures (Gallo et al., 2007). Error rates were lower when study format was mutually exclusive, indicating higher occurrence of recall-to-reject. However, it cannot be deduced whether mutual exclusivity is a necessary precondition or not.

Brainerd and colleagues (2003) assume that if study lists contain associated items, recall-to-reject can still be used intuitively on the basis of a feeling that the associated distractors' familiarity is satisfactorily accounted for by the recollected target. Similarly, Gallo (2004) assumes that recall-to-reject can also be used if stimuli are not mutually exclusive. However, he argues that this requires that participants are aware of the number of associated items and are able to exhaustively recall them. Thus, Gallo supposes that recall-to-reject is a strategic process that implies the logical exclusion of an associated distractor based on the study list structure. In his study, associated items were exemplars of the same semantic category (e.g., parsley, rosemary, thyme). Consistent with his assumptions, Gallo found stronger evidence for recall-to-reject in a constant length condition compared to a varied length condition, and only if the number of category exemplars presented during study was small (three compared to five items). Furthermore, the false recognition effect was lower when presented exemplars of a category were exhaustively recalled. These results are in line with the study of Tussing and Greene (1999), where category exemplars were also used as associated stimuli. When a distractor was associated to five similar items presented during study, a decrease of the false alarm rate was found when items were repeated five or ten times, but only if the items were presented in blocked format. Presenting the items in blocked format possibly increased participants' awareness of how many exemplars were presented per category as well as the recall of items of a specific category. However, please note that the results of Tussing and Greene could alternatively be explained with the memory editing mechanism called *identification*. Extensive repetition of items during study might have led participants to think of other associated stimuli and became aware that these were not presented during study.

Moreover, the results of the judgments of frequency study that was presented as indirect evidence for recall-to-reject (C. M. Jones & Heit, 1993), can be interpreted in favour of the hypothesis that mutual exclusivity is a necessary precondition for recall-to-reject. Remember that repeated presentation of a single similar word affected the frequency judgment only of presented but not of nonpresented words. The lack of an effect for nonpresented items violated the total similarity principle of global memory models and was interpreted as indirect evidence for recall-to-reject. The difference can be explained by the consideration that stimuli are mutually exclusive in the nonpresented condition but not in the presented condition. To continue, stimuli in the nonpresented condition in which participants are asked to judge the frequency of an associated distractor, are no longer mutually exclusive when multiple similar items are presented during study instead of a single similar item. Therefore, it can be hypothesised that the number of similar items presented during study should influence recognition of an associated distractor. In line with this hypothesis is the finding that the false alarm rate of an associated distractor was not increased when the corresponding presented similar item was repeated five times during study, but when five different similar words were presented during study (Tussing & Greene, 1999).

Overall, evidence suggests that recall-to-reject is a rather strategic process that either requires the mutual exclusivity of stimuli or the exhaustive recall and awareness of the number of associated items presented. One goal of the present work was to further investigate the importance of mutual exclusivity of stimuli.

5.5.3 Type of Associated Distractor

As the recall-to-reject process is based on the recall of the corresponding studied item, the probability of recall-to-reject should increase with the tendency of the associated distractor to trigger recall of the corresponding studied item. This assumption could explain why evidence for recall-to-reject has often been found in associative but not in item recognition (Rotello & Heit, 2000). For example if A-B and C-D are word pairs presented during study and A-D is a rearranged pair presented as associated distractor during test, then A can be used as a retrieval cue to recall that B was studied with A and not D.

The assumptions about what type of associated distractor is a good retrieval cue for its corresponding target differ though. The importance of perceptual similarity between studied items and their corresponding associated distractors has been emphasised (Brainerd et al., 2006; Brainerd et al., 1995) as well as the importance of conceptual overlap or semantic similarity (Brainerd et al., 2003; T. C. Jones & Jacoby, 2001; Odegard et al., 2005).

It has been argued that recall-to-reject may have a limited effect in reducing memory conjunction errors because conjunction lures differ conceptually from their parent words (T. C. Jones & Jacoby, 2001). For instance, the conjunction lure *blackbird* is conceptually different from the parent words *blackmail* and *jailbird*. Odegard and colleagues (2005) varied the semantic similarity shared between parent words and their corresponding conjunction and feature lures. For instance, when *handball* and *shotgun* are presented as parent words, the conjunction lure *handgun* is semantically very similar to its parent *shotgun*. The results are ambiguous, the amount of semantic similarity did not have any effect on the false alarm rate but more self reported recall-to-reject and more high confidence rejections were found in the high similarity condition.

Evidence for the importance of the conceptual relationship between cue and target stems from a discrimination task in which participants had to discriminate between read and self generated words (T. C. Jones, 2006). Participants were required during study to read words and additionally to generate words, either with an anagram procedure (that requires mentally switching the locations of two misplaced letters, e.g., fofcee for coffee) or a conceptual cue procedure (e.g., black, caffeinated beverage: co___). Results show that participants were better able to discriminate between read and self generated words in the read-conceptual cue procedure than in the read-anagram procedure, which emphasises the role of conceptual information.

The finding that the recall-to-reject parameter of the conjoint recognition model was higher for synonyms as opposed to antonyms supports the importance of semantic similarity (Brainerd et al., 1999). Antonyms versus rhymes have also been used as associated distractors to manipulate the type of similarity (Odegard & Lampinen, 2005). Conjoint recognition parameter for recall-to-reject was greater when associated distractors were antonyms as opposed to rhymes of previously studied items. In an additional experiment, self reports for recall-to-reject were more frequent for antonyms compared to rhymes, but the false recognition effect was the same for both types of associated distractors. Similarly, mixed evidence for the hypothesis about the importance of surface details was found by Brainerd and colleagues (2006).

In summary, the probability of recall-to-reject seems to depend on the ability of the associated distractor to trigger the recall of the corresponding target. This can explain the failure of many studies to provide evidence for recall-to-reject in item recognition compared to associative recognition. Furthermore, recall-to-reject seems to be more likely if a distractor is semantically associated to the corresponding target compared to perceptually associated. One goal of the present work was to provide evidence for recall-to-reject where it has been very rare to date, that is, in item recognition tasks.

6 Current State of Research

Recall-to-reject is one of the central memory editing mechanisms that can oppose high familiarity and reduce the occurrence of false memories. Evidence for recall-to-reject is in line with dual-process theories, which state that recognition decisions are based on familiarity as well as on recollection processes. However, due to the diversity of tasks and methods that have been used to investigate recall-to-reject, research to date has yielded many inconsistent findings. Overall, evidence for recall-to-reject is very rare for item recognition compared to associative recognition and discrimination tasks. Most noteworthy is the fact that, although manipulations of recall-to-reject are predicted to have an effect on the false recognition effect, most studies failed to find evidence for recall-to-reject from recognition data. Instead, most evidence for recall-to-reject is limited to model parameters. The reasons for this discrepancy between recognition data and model parameters, which has important theoretical and practical implications, remain unstudied. Additionally, the role of moderating factors such as explicit instructions or mutual exclusivity of stimuli has not been clarified yet.

The present work aimed at providing evidence for the hypothesis that recall-to-reject can be used to reduce the false recognition effect. A simple item recognition test and analysis of recognition data was chosen as previous research lacks evidence for recall-to-reject in this domain. A category cueing manipulation that has been shown to improve recall but failed to date to affect recognition was used in all experiments to increase the probability of recall-to-reject. Experiment 1 tested the hypothesis that category cueing can increase recall-to-reject and consequently reduce the false recognition effect even in an item recognition task and without explicit instructions. In contrast, explicit recall-to-reject instructions were given in Experiment 2. The goal of Experiment 3 was to eliminate the side effects of category cueing that were observed in the first two experiments. Finally, Experiment 4 was designed to provide further information about the necessity of mutual exclusivity of stimuli for the effectiveness of recall-to-reject.

7 Experiment 1

The purpose of the first experiment was to investigate whether category names, when presented as retrieval cues, can be used to reduce false recognition via facilitating recall-to-reject. An item recognition task with a semantic similarity manipulation was developed. Associated distractors were exemplars from a category of which another exemplar had already been presented during study. Associated stimuli were mutually exclusive, that is, only one exemplar per category was presented during study to maximise the effectiveness of the recall-to-reject process. During test category names were presented prior to the items and in the form of questions to which an old/new recognition decision was required. For instance, participants were asked if a word from the category fruit had been presented during study. This was done to guarantee that participants were attending to the retrieval cues and to maximise their effectiveness. Participants were not given explicit recall-to-reject instructions.

In a non-cued control condition, a typical false recognition effect was expected: The false alarm rate for associated distractors was anticipated to be significantly higher than for standard distractors. It was assumed that category cueing would reduce the false recognition effect expected for non-cued items. If category names are successful in facilitating recall-to-reject, the false recognition effect should be smaller to the point of being eliminated for cued items as compared to non-cued items.

7.1 Method

7.1.1 Participants

Participants were 74 adults (mostly students), 46 of whom were female. The mean age of participants was 25.96 (*SD* = 6.36), ranging from 17 to 50 years. All of them were native German speakers. Participants were tested in groups of up to four and were paid € 5.00 for their participation.

7.1.2 Materials

The item material consisted of 108 category names with two category exemplars each (e.g., fruits: apple, pear). All category exemplars were German nouns with one to five syllables. Category exemplars were matched for frequency according to a German vocabulary online encyclopaedia (Wortschatz Lexikon Deutsch, 2005), word length in number of syllables (mean difference *M* = 0.46, *SD* = 0.60), dominance (for those categories where norms are available, see Mannhaupt, 1983; Scheithe & Bäuml, 1995), and orthographical similarity. The item set did not contain any compound words or homonyms (i.e., words with more than one meaning), and the categories did not overlap (i.e., exemplars did not belong to more than one category). See Appendix A for a complete list of the item material used in Experiment 1.

Participants studied category exemplars and were later tested for recognition memory. Only one of the two exemplars of a category was presented in the study phase. Test items that were presented in the study phase are termed targets. Test items are termed standard distractors if neither the word itself nor the other exemplar of the same category was presented in the study list. Associated distractors were test items that were not presented in the study list whereas the other exemplar of the same category was presented in the study list (e.g., *pear* was learned and *apple* was presented as the associated distractor).

In the test phase all items were preceded by a category question: A category name was presented and participants had to decide whether an exemplar of this category had been in the study list or not. An example of all item types is shown in Table 2. For all *cued* items the preceding category question was about the corresponding category (e.g., if *pear* was learned and *apple* was presented as the associated distractor, the preceding category question asked whether a *fruit* was learned). In order to have an appropriate control condition, all *non-cued* items were also preceded by a category question, but this was about some unrelated category (e.g., the category question of the example above could ask whether a *beverage* was learned)[5].

[5] For non-cued items, it was varied whether an exemplar of the preceding unrelated category question had been learned or not. Additionally presented exemplars of unrelated categories were randomly chosen from the item material.

Table 2

Examples for the different item conditions in Experiment 1. All non-cued items are preceded by an unrelated category question, whereas the category question preceding cued items was about the category corresponding to the item tested. For simplicity, the table does not show items of the study list referring to the category question of non-cued items.

	Study list	Test list	
Item condition	Item referring to item question	Category question	Item question
Non-cued targets	apple	beverage?	apple?
Non-cued distractors	–	beverage?	apple?
Non-cued associated distractors	pear	beverage?	apple?
Cued targets	apple	fruit?	apple?
Cued distractors	–	fruit?	apple?
Cued associated distractors	pear	fruit?	apple?

Categories were randomly assigned to item types for each participant. For targets as well as standard distractors it was decided at random which exemplar of a category was to be used as target and distractor, respectively. For associated distractors, it was also randomly decided which of the two exemplars was presented in the study list and which was presented as the associated distractor in the test list.

Participants studied three lists, each consisting of a total of 32 items. All words in the study list were presented in random order except for the first and last five items, which were primacy and recency buffers that were not tested and therefore not included in the analysis.

Each of the three test lists consisted of 24 items preceded by their respective category question. All items of a test list were presented in random order. Each test list was composed of eight targets, distractors, and associated distractors respectively, half of which were cued and half were uncued.

7.1.3 Procedure

First, participants were familiarised with the recognition task. They had to work through at least one practice trial with a short study and test list. The test list of the practice phase comprised one test item of each item condition (compare Table 2) preceded by the corresponding category question. Participants were free to reread instructions and complete another practice trial or to start the experiment. In the subsequent study phase, words were presented on a computer monitor for 1000 ms each, with a blank screen shown for 300 ms in the interval between two word presentations. Participants were only instructed to memorise words as well as possible. They were not informed about details of study or test list construction.

A 30 sec interval filled with simple arithmetic problems separated study and test phase. Participants had to decide whether the solution of presented arithmetic tasks (e.g., 34 + 21 = 65) was correct or not. Participants were instructed to finish as many tasks as possible while keeping errors to a minimum. At the end of the distractor task participants received a feedback about the number of arithmetic tasks completed and the number of tasks rated correctly.

In the test phase, category and item questions were presented alternately. Within category questions, a category name was presented and participants' task was to indicate whether an exemplar of this category had been presented in the study phase. For the following item recognition task, a category exemplar was presented and participants had to decide whether this exemplar had been in the study list or not. Display examples for the presentation of a category and item question are provided in Figure 4. Each display of the category questions consisted of the question *Was a word of the following category in the study list?* presented at the top of the display and the category name presented at the centre of the screen. Each display of the item recognition task consisted of the item question *Was the following word presented in the study list?* presented at the top of the display and the item presented at the centre of the screen[6].

[6] To ensure an easy differentiation between category question and item recognition task, the category name and the word *category* within the category question were in bold font and coloured orange whereas the item and the word *word* within the item recognition question were in bold font and coloured black.

Figure 4 Display examples for the presentation of a category and item question of the recognition test. The word *Kategorie [category]* and the category name (here *Obst [fruit]* were orange coloured. Corresponding to the colour of the response buttons on the keyboard, the she small boxes above the response options were red for *nein [no]* and blue for *ja [yes]*.

Participants responded by pressing the *yes* (right arrow) or *no* (left arrow) key on a keyboard with their index fingers. The recognition task was self-paced. After practice trials, each participant was required to complete three blocks, each consisting of study phase, retention interval, and test phase. Blocks were separated by a one minute break without a distractor task. After each block, participants received an overall feedback regarding the percentage of category and word recognition questions answered correctly. After the last block, participants were informed about the purpose of the experiment upon request. The experiment lasted approximately 30 minutes.

7.1.4 Design

Experiment 1 employed a two-factorial design with item type (target, distractor, associated distractor) and category question (related = cued vs. unrelated = non-cued) as within-subject variables. This resulted in six different item conditions which are termed cued/non-cued targets, cued/non-cued distractors, and cued/non-cued associated distractors. Recognition performance was the dependent variable. The relative frequency of an old response in the old/new recognition test was registered for the different item conditions.

The comparison between the false alarm rate for associated distractors and for standard distractors defined the false recognition effect. A power analysis using G*Power 3 (Faul, Erdfelder, Lang, & Buchner, 2007) was conducted to determine the probability of observing a significant false recognition effect in the present study. Given $N = 74$ participants and $\alpha = .05$, it was possible to detect a false recognition effect of medium size $d_z = 0.5$ (according to the effect size conventions of Cohen, 1988) with a probability of $1 - \beta = .99$. Please note that all power analyses reported in the present work were conducted using G*Power 3.

As hypotheses refer to binary comparisons of recognition performance between specific item conditions, analyses are based on paired sample *t*-tests rather than a conventional multivariate analysis of variance[7]. This applies to all experiments reported in the present work. To protect against α-error accumulation, the error probability level of $\alpha = .05$ was corrected according to the Bonferroni-Holm method (Holm, 1979) for all statistical tests performed within the unconditional and conditional analysis, respectively (see Chapter 7.2 for details). This applies to all experiments of the present work. For all statistical analyses reported in the present work, d_z is reported as a measure of the sample effect size (Cohen, 1988).

The dependent variable *relative frequency of old responses* is based on binary data (the answer to the question whether an item has been in the study list or not is either yes or no). When analyzing proportions it has to be taken into account that the distribution is only approximately normal and variances depend on its particular mean (Winer, 1971). Therefore, violations of the assumption of equality of variance and the assumption of normal distribution of data challenge the use of conventional statistical tests such as *t*-tests or analyses of variance. However, arcsine transformation of proportions can be used to improve the equality of variances and normal distribution (Cohen, 1988; Winer, 1971). Furthermore, it has been shown that *t*-tests are robust and violations of the normality assumption have little effect upon α-error probability (Myers & Well, 1995). With respect to the present experiment, analyses of arcsine-transformed propor-

[7] The main part of a conventional multivariate analysis of variance would not be of interest anyway since, for example, a main effect of the factor item type is trivial, as the probability of an old response will definitely differ between targets and distractors. Hypotheses could only be tested by post-hoc analyses anyway.

tions[8] resulted in identical results and conclusions, as did those analyses performed on non-transformed proportions. For simplicity, the results of statistical analyses performed on untransformed data are reported for Experiment 1 as well as the following experiments.

7.2 Results

Participants answered 71% of the category questions (*SE* = .011) and 86% of the item questions (*SE* = .009) correctly. The relative frequencies of old responses for the different item conditions are shown in Table 3.

In the non-cued condition[9], the false alarm rate for associated distractors was significantly higher than for standard distractors, that is, a typical false recognition effect was observed ($t(73) = 6.15$, $p < .001$, $d_z = 0.71$). In the cued condition, the false alarm rate for associated distractors and standard distractors did not differ ($t(73) = 0.41$, $p = .683$, $d_z = 0.05$), that is, category cueing eliminated the false recognition effect. However, a look at the absolute false alarm rates reveals that the false alarm rate for cued standard distractors was higher than for non-cued standard distractors ($t(73) = 4.68$, $p < .001$, $d_z = 0.54$) and the false alarm rate for associated distractors did not differ between cued and non-cued items ($t(73)=0.45$, $p = .654$, $d_z = 0.05$). The hit rate for cued targets was higher than that for non-cued targets ($t(73) = 2.43$, $p = .009$, $d_z = 0.28$). All statistical analyses reported so far are based on an error probability level of $\alpha = .05$ corrected according to the Bonferroni-Holm method (Holm, 1979) to protect against α-error

---

[8] Arcsine transformation was performed using the following formulas: $x' = 2\arcsin\sqrt{x}$ if x > 0 and x < 1, $x' = 2\arcsin\sqrt{0+\left[1/(2n)\right]}$ if x = 0 and $x' = 2\arcsin\sqrt{1-\left[1/(2n)\right]}$ if x = 1, where n refers to the number of observations x is based on (Winer, 1971).

[9] Preliminary analysis confirmed that the relative frequencies of old responses for non-cued targets, distractors and associated distractors did not differ depending on whether an exemplar of the preceding unrelated category had been learned or not ($t(73) = 1.79$, $p = .077$, $d_z = 0.21$ for targets, $t(73) = 0.95$, $p = .346$, $d_z = 0.11$ for distractors, and $t(73) = 0.09$, $p = .925$, $d_z = 0.01$ for associated distractors). All subsequent analyses on non-cued items could therefore be performed on collapsed data.

accumulation. However, the correction of the error probability level did not change the significance of results.

Table 3

Proportion of hits and false alarms for different item conditions in Experiment 1 (standard errors in parentheses). The table shows relative frequencies of old responses both in overall terms and depending on the answer to the preceding category question. For the conditional analysis[10], *t*-values, degrees of freedom (*df*) and effect sizes (d_z) are also listed.

Item condition	Unconditional	Conditional			
		Preceding category accepted	Preceding category rejected		
	M (*SE*)	*M* (*SE*)	*M* (*SE*)	*t* (*df*)	d_z
Non-cued targets	.797 (.020)	.807 (.023)	.800 (.027)	0.07 (71)	0.01
Non-cued distractors	.050 (.009)	.036 (.010)	.059 (.014)	1.68 (71)	0.20
Non-cued associated distractors	.136 (.013)	.153 (.019)	.141 (.021)	0.36 (72)	0.04
Cued targets	.840 (.020)	.923 (.017)	.629 (.041)	7.26 (64) *	0.91
Cued distractors	.137 (.020)	.266 (.037)	.072 (.016)	5.91 (68) *	0.72
Cued associated distractors	.144 (.017)	.152 (.020)	.120 (.027)	0.89 (65)	0.11

* $p < .001$

Recognition data were additionally analysed depending on the answer to the preceding category question (conditional analysis). As preceding category questions for non-cued items are unrelated to the following item, acceptance or rejection of the preceding category question was not expected to have any effect on the following item recognition decision. However, as preceding category questions for cued items are semantically related to the following item, item recognition decisions for cued items are likely to be related to the preceding category question. For cued targets the hit rate was expected to be higher if the preceding category question had been accepted. This was

[10] As can be seen in Table 3, sometimes both probabilities of the conditional analysis were higher than the unconditional probability. This might initially seem unusual but is nevertheless correct. The effect is due to the varying number of trials on which the conditional probabilities are based. This effect cannot be prevented since the conditional analysis is based on participants' answer to the preceding category question. Weighing probabilities with the corresponding number of trials would attach different values to participants and would therefore be inadvisable.

expected since the probability of recognizing a target item should be very high after the corresponding category question had been accepted based on either recall or familiarity. For cued associated distractors, the false alarm rate was expected to be lower if the preceding category question had been accepted. This was thought to result from recall-to-reject because if the preceding category question has been accepted based on recall of the corresponding studied category exemplar, the discrepancy to the following associated distractor is likely to be detected and the associated distractor will be rejected. For cued standard distractors, the conditional false alarm rates were expected not to differ because acceptance of a category question cueing a distractor was assumed to be based on guessing. All statistical tests within the conditional analysis were based on an error probability level of $\alpha = .05$ corrected according to the Bonferroni-Holm method (Holm, 1979).

As can be seen in Table 3, the answer to an unrelated category question preceding non-cued items did not have any effect on the probability of an old response for all item types. However, the hit rate for cued targets and the false alarm rate for cued standard distractors was significantly higher when the preceding category question had been answered with yes. The false alarm rate for cued associated distractors was not affected by the answer to the preceding category question.

7.3 Discussion

Cueing items by the corresponding category name did not only reduce but even eliminate the false recognition effect. This indicates that category cueing was successful in increasing the probability of the recall-to-reject process by facilitating conscious recollection of learned items. This interpretation is compatible with the higher hit rate for cued compared to non-cued targets. Thus, the present experiment confirms the assumption that participants are capable of identifying the test structure themselves and spontaneously use recall-to-reject. Thus, explicit recall-to-reject instructions are not necessary to find evidence of recall-to-reject.

However, the absolute false alarm rate for cued associated distractors was not reduced compared to that for non-cued associated distractors. It is assumed that this is due to a general increase in familiarity caused by the corresponding category questions. It seems likely that the semantic similarity between category names and corresponding exemplars increases the familiarity of exemplars in general. This could occur for example through a process of spreading activation (Collins & Loftus, 1975), or a generate-recognize strategy (Anderson & Bower, 1972; Bahrick, 1970), a better accessibility of the gist trace in the context of the fuzzy trace theory (e.g., Brainerd & Reyna, 2002a), or it is possible that category and exemplar constitute a compound cue (Ratcliff & McKoon, 1988). This side effect of category cueing should affect all cued item types and result in an increased probability of an old response. The results show that this was true for standard distractors and targets (although the higher hit rate for cued compared to non-cued targets could be due to both an increase in familiarity and conscious recollection). However, the false alarm rate for cued associated distractors was not increased compared to non-cued associated distractors. This indicates the occurrence of the recall-to-reject process, which reduces the false alarm rate. Since both processes–recall-to-reject and familiarity–work in opposition, the absolute false alarm rate for associated distractors is neither increased nor reduced.

The conditional analysis, that is, the analyses of the responses to the item questions depending on the answer to the preceding category question, shows that, as expected, it did not make a difference for all non-cued item types whether the preceding unrelated category question had been accepted or rejected. This was expected since the preceding category questions for non-cued items were unrelated to the following item. In contrast, the hit rate for cued targets and the false alarm rate for cued standard distractors were higher if the preceding category question had been accepted. This difference found for cued targets was expected due to recall or familiarity processes. The same difference found for cued distractors was not expected and can only be based on guessing. This indicates that participants tended to respond such that their answer to the item question was consistent with their answer to the preceding corresponding category question. Such a response strategy should affect all cued item types, but no difference was found for cued associated distractors. In fact, an opposite pattern was expected due to the recall-to-reject process. Recall-to-reject is most likely to occur if the preceding category question has been accepted based on recall of the corresponding category exemplar. Therefore, recall-to-reject should lower the false alarm rate after accepted compared to rejected category questions. But if participants do employ a strategy of answering consistently the opposite pattern is expected. Hence, both processes work in opposition and seem to cancel each other out.

It should be pointed out that the tendency to answer consistent with the preceding category question becomes obvious in the conditional analysis but nevertheless influences the unconditional data, too. However, it is difficult to make precise predictions about the exact effect on the unconditional data. Whether the increase of item acceptance after accepted category questions will be stronger than the decrease of item acceptance after rejected category questions will depend on the probability of category acceptance as well as the initial probabilities of item acceptance without the influence of the response strategy. For cued standard distractors, the probability of rejecting the preceding category will be higher than that for accepting it (as no exemplar of the category has been learned). As the initial probability of rejecting a standard distractor should be higher than that of accepting it (as the item has not been studied), the tendency to answer consistently should result in a decrease of false alarms for standard distractors. By contrast, for associated distractors, the probability of accepting the preceding category will be higher than rejecting it (as an exemplar of the category has been learned) and as the initial probability of rejecting an associated distractor should be higher than accepting it, the tendency to answer consistently should result in an increase of false alarms. Both effects, a decrease of false alarms for standard distractors and an increase of false alarms for associated distractors, increase the false recognition effect and therefore work in opposition to the recall-to-reject process.

The hypothesised effects of a strategy to answer consistently on the false alarm rates can be retraced in Figure 5, which illustrates the relation between the conditional probabilities and the overall false alarm rate. The left column illustrates the response probabilities for cued standard distractors as well as for cued associated distractors estimated from the observed data. The right column illustrates the hypothetical response probabilities that should occur without the influence of a strategy to answer consistently (hypothetical values in grey rectangles). On the premise that the acceptance of a category question preceding a standard distractor is based on guessing, it was hypothesised that the probability of false acceptance of standard distractors does not depend on the answer to the preceding category question (compare Figure 5, .20 after accepted as well as after rejected category question). Assuming that recall-to-reject does occur for associated distractors, the probability of false acceptance was hypothesised to be lower after acceptance of the preceding category question (.05) compared to a rejection (.20).

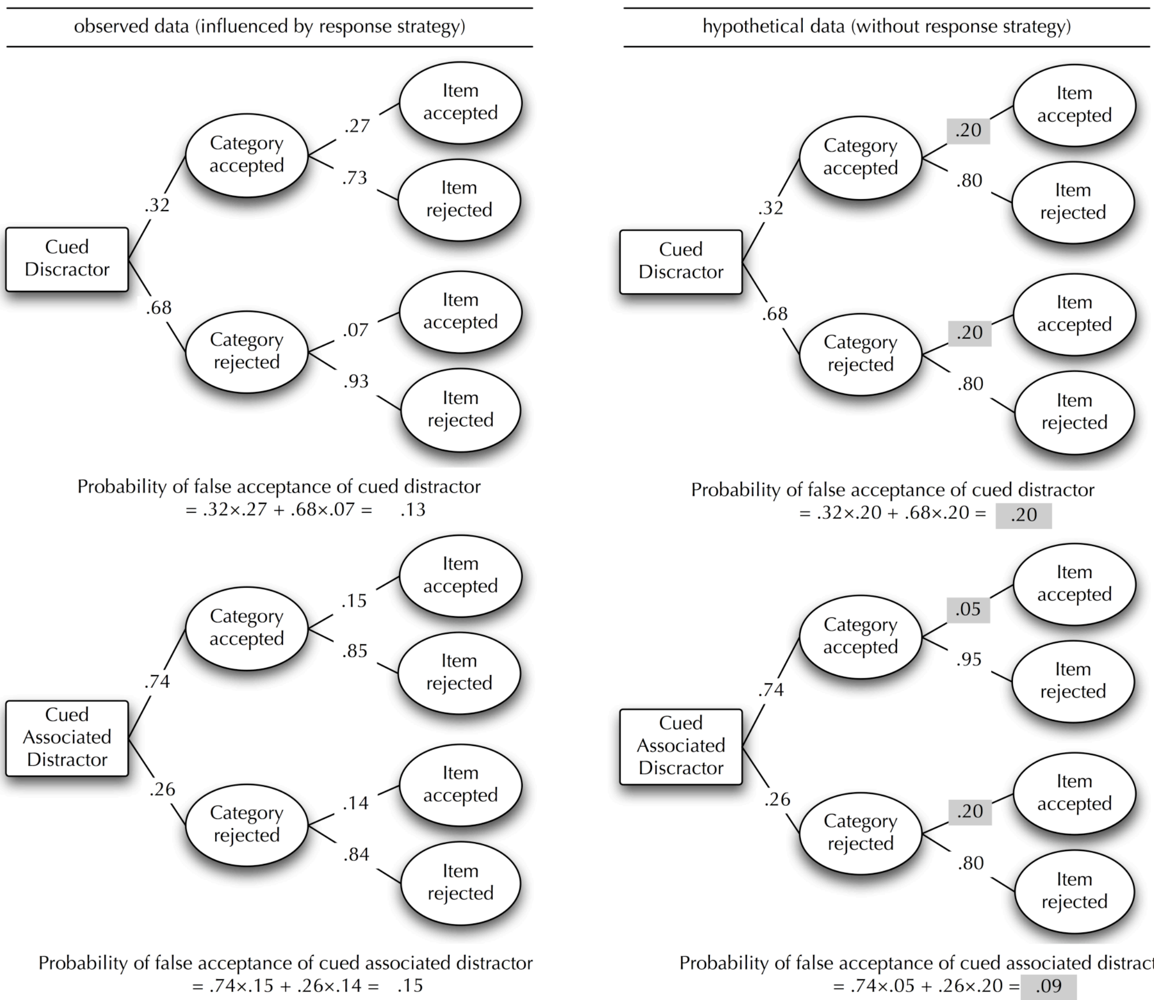

Figure 5 Illustrative example for the influence of the strategy to answer in a manner consistent with the preceding category question on the overall false alarm rates. The figure presents the possible answers to both types of distractors and their preceding category questions. It is assumed that the relative frequencies depicted on the left hand side are influenced by the response strategy whereas the hypothetical probabilities depicted on the right hand side are not. The example illustrates that the response strategy results in a decrease of false alarms to standard distractors and in an increase of false alarms to associated distractors.

The hypothetical values were chosen to approximate the observed data but nevertheless to illustrate that the observed data reflect the influence of a strategy to answer consistently whereas the hypothetical probabilities do not. Thus, the hypothetical probability of accepting an item after an accepted category question had to be lower compared to the corresponding observed relative frequency (compare Figure 5, .20 vs. .27 for standard distractors and .05 vs. .15 for associated distractors, respectively). Accordingly, the hypothetical probability of rejecting an item after a rejected category question had to be lower compared to the corresponding observed relative frequency (compare Figure 5, .80 vs .93 for standard distractors and .80 vs. .84 for associated distractors, respectively). Consistent with the predictions above, for standard distractors, the observed overall false alarm rate was lower compared to the hypothetical overall false alarm rate (.13 vs. .20). Similarly, for associated distractors, the observed overall false alarm rate was higher compared to the hypothetical overall false alarm rate (.15 vs. .09). Thus, the contribution of the strategy to answer consistently seems to increase the false alarm rate for cued associated distractors and decrease the false alarm rate for cued standard distractors. The resulting increase of the false recognition effect opposes the recall-to-reject process. However, please note that hypothetical false alarm rates cannot be compared between standard and associated distractors (.20 vs. .09 in Figure 5) because conditional probabilities are only hypothetical.

To summarise, category names used as retrieval cues are successful in increasing the probability of the recall-to-reject process without explicit recall-to-reject instructions. Category cueing can even eliminate the false recognition effect found for non-cued items. However, the semantic similarity between category names and corresponding exemplars simultaneously increases the familiarity of cued exemplars in general. Due to this side effect the baseline of false alarms is higher compared to non-cued items and therefore no absolute reduction of the false alarm rate can be achieved. Furthermore, the obvious semantic relation between category question and the following category exemplar encourages a tendency of answering consistently. This second side effect is assumed to increase the false alarm rate for associated distractors and therefore works in opposition to the recall-to-reject process. Concurrently, the tendency to answer consistently is assumed to decrease the false alarm rate for standard distractors, which should further increase the false recognition effect and therefore also works in opposition to the recall-to-reject process.

8 Experiment 2

Experiment 2 was designed to replicate the results of Experiment 1 with a slightly modified procedure. The unrelated control condition of Experiment 1 was replaced by an item recognition task without any preceding category question. Only for the cued conditions category questions were presented before the items. These category questions were always related to the subsequent item. This was done to assure that the cue (i.e., the category question) was perfectly valid and participants therefore made a maximum effort to recall learned items. Furthermore this manipulation might have an effect on the attribution of familiarity induced by the relatedness of the category question and the following item. As category questions are always followed by a corresponding category exemplar, participants could expect cued items to be somewhat familiar. This might help participants to attribute, at least partly, the familiarity to the conceptual overlap between category and item question, which would prevent an influence on recognition judgments (for the role of attributional processes, see Lampinen et al., 2005; McDermott & Watson, 2001; Whittlesea & Leboe, 2003). Additionally, explicit recall-to-reject instructions were given to maximise the utilisation of the recall-to-reject process.

8.1 Method

The Design of Experiment 2 was nearly identical to that of Experiment 1. Thus, only differences are described.

8.1.1 Participants

Participants were 75 students, 45 of whom were female. The mean age of participants was 26.01 (*SD* = 6.57), ranging from 19 to 47 years. All of them were native German speakers. None of them had participated in Experiment 1. Participants were tested in groups of up to four and were not paid for participation since psychology students conducted participant recruitment as part of the requirements for obtaining course credit.

8.1.2 Materials

The same item material as in Experiment 1 was used. As in Experiment 1, participants studied category exemplars and were later tested for recognition memory. There were the same item types: targets, distractors, and associated distractors.

In contrast to Experiment 1, in the test phase only cued test items were preceded by a category question. In the control condition of Experiment 2, non-cued items were not preceded by a category question. Therefore, category questions always corresponded to the following item.

Participants studied three lists consisting of a total of 34 items. All words in the study list were presented in random order except for the first and last five items, which were primacy and recency buffers that were not tested and therefore not included in the analysis.

Each of the three test lists consisted of 36 items half of which were preceded by a category question. Each test list was composed of twelve targets, distractors and associated distractors respectively, half of which were cued and half were non-cued.

8.1.3 Procedure

Concerning the procedure, the only difference between Experiment 2 and Experiment 1 was that participants were informed that only one exemplar of a category was presented in the study list (i.e., category exemplars were mutually exclusive). Instructions emphasised that if participants were able to recall one category exemplar they could be sure that other category exemplars presented in the test list were new and therefore should be rejected.

8.1.4 Design

Experiment 2 employed a two-factorial design with item type (target, distractor, associated distractor) and category question (with = cued vs. without = non-cued) as within-subject variables. Similarly to Experiment 1, this resulted in six different item conditions, namely cued/non-cued targets, cued/non-cued distractors and cued/non-cued associated distractors. Recognition performance was the dependent variable.

The comparison between the false alarm rate for associated distractors and for standard distractors defined the false recognition effect. Power analysis showed that given $N = 75$ participants and $\alpha = .05$, it was possible to detect a false recognition effect of medium size $d_z = 0.5$ with a probability of $1 - \beta = .99$.

8.2 Results

Participants answered 69% of the category questions (*SE* = .010) and 85% of the item questions (*SE* = .008) correctly. The relative frequencies of old responses for the different item conditions are shown in Table 4.

Table 4

Proportion of hits and false alarms for different item conditions in Experiment 2 (standard errors in parentheses). The table shows relative frequencies of old responses both in overall terms and depending on the answer to the preceding category question. For the conditional analysis, *t*-values, degrees of freedom (*df*) and effect sizes (d_z) are also listed.

Item condition	Unconditional	Conditional			
		Preceding category accepted	Preceding category rejected		
	M (*SE*)	*M* (*SE*)	*M* (*SE*)	*t* (*df*)	d_z
Non-cued targets	.814 (.015)				
Non-cued distractors	.076 (.009)				
Non-cued associated distractors	.144 (.015)				
Cued targets	.812 (.017)	.892 (.015)	.579 (.038)	8.59 (69) *	1.03
Cued distractors	.156 (.016)	.291 (.029)	.057 (.012)	7.85 (71) *	0.93
Cued associated distractors	.144 (.015)	.151 (.019)	.139 (.028)	0.48 (68)	0.06

* $p < .001$

The overall pattern of results corresponded to Experiment 1. In the non-cued condition, the false alarm rate for associated distractors was significantly higher than for standard distractors, that is, a typical false recognition effect was observed ($t(74) = 4.56$, $p < .001$, $d_z = 0.53$). In the cued condition, the false alarm rate for associated distractors and standard distractors did not differ ($t(74) = 0.79$, $p = .433$, $d_z = 0.09$), that is, category cueing eliminated the false recognition effect. Similar to Experiment 1, a look at the absolute false alarm rates reveals that the false alarm rate for cued standard distractors was higher than for non-cued standard distractors ($t(74) = 5.18$, $p < .001$,

d_z = 0.61) and the false alarm rate for associated distractors did not differ between cued and non-cued items (t(74)= .08, p = .936, d_z = 0.01). However, in contrast to Experiment 1, the hit rate for cued targets was not higher than that for non-cued targets (t(74) = 0.13, p = .447, d_z = 0.02).

As items of the control condition in Experiment 2 were not preceded by any category question, recognition data depending on the answer to the preceding category question (conditional analysis) were only available for cued items (see Table 4). Results of the conditional analysis correspond to Experiment 1. The hit rate for cued targets and the false alarm rate for cued standard distractors was significantly higher if the preceding category question had been answered with yes. The false alarm rate for cued associated distractors was not affected by the answer to the preceding category question.

8.3 Discussion

The overall pattern of results is consistent with Experiment 1. Cueing items by the corresponding category name did eliminate the false recognition effect. Replicating this main effect of category cueing strengthens the interpretation that category cueing can be used to increase the probability of the recall-to-reject process. The procedure of Experiment 2 was slightly modified compared to Experiment 1 in order to maximise the utilisation of the recall-to-reject process. Firstly, the unrelated control condition of Experiment 1 was replaced by an item recognition task without any preceding category question. Consequently, in Experiment 2 category questions were always followed by a corresponding category exemplar and therefore category questions were always valid and participants should make a maximum effort to recall learned items. Secondly, explicit recall-to-reject instructions were given to maximise the utilisation of the recall-to-reject process. The pattern of results was the same for both experiments and a comparison of the false alarm rate between standard distractors and associated distractors for cued items reveals no significant differences (at a descriptive level the false alarm rate for cued associated distractors is slightly lower compared to cued standard distractors in Experiment 2 but this difference is only minimal and far from statistical significance). However, for non-cued items, the size of the false recognition effect was somewhat larger in Experiment 1 (d_z = 0.71 which is between a medium and a large effect according to the effect size conventions of Cohen, 1988) than in Experiment 2 (d_z = 0.53 which corresponds to a medium effect). This suggests that the probability of spontaneous recall-to-reject might have been higher in Experiment 2 than in Experiment 1. This difference is likely to be due to the recall-to-reject instructions because the replacement of the control condition should only have an effect for cued items.

Beside the elimination of the false recognition effect, however, category cueing again increased the level of false alarms for standard distractors, possibly due to an increase of familiarity caused by the semantic similarity between category names and corresponding exemplars. A look at the effect sizes for this increase of the false alarm rate for cued compared to non-cued standard distractors (d_z = 0.54 in Experiment 1 vs. d_z = 0.61 in Experiment 2) indicates that the replacement of the control condition did not alter this side effect of category cueing. As category questions in Experiment 2 were always followed by a corresponding category exemplar, participants were assumed to expect cued items to be somewhat familiar. This might have helped participants to correctly attribute the additional familiarity to the conceptual overlap between category and item question and therefore might have prevented to influence recognition judgments. However, this hypothesis could not be confirmed.

The results of the conditional analysis are consistent with Experiment 1. This replication strengthens the interpretation that the expectation for cued associated distractors (i.e., lower false alarm rate after accepted category question) was not fulfilled due to a second side effect of category cueing. As recall-to-reject is most likely to occur if the preceding category question has been accepted based on recall of the corresponding category exemplar, a lower false alarm rate after accepted compared to rejected category questions was expected. However, the observed higher false alarm rate for cued standard distractors after accepted compared to rejected category questions indicates that participants again tended to respond such that their answer to the item question was consistent with their answer to the preceding corresponding category question. This tendency works in opposition to recall-to-reject and both processes seem to cancel each other out.

Furthermore, the tendency to answer consistently can explain the sole unexpected result concerning target items. Although the hit rate of cued targets was expected to be higher than that for non-cued targets due to an increase in conscious recollection and/or familiarity, no such increase was found in Experiment 2. Of course this could be due to a ceiling effect, but could alternatively be caused by the tendency to answer consistently. Considerations that have been made with respect to the effect of response strategies in the Discussion of Experiment 1 lead to the assumption that a strategy to answer consistently should result in a decrease of the hit rate for targets. Hence, the tendency of answering consistently works against the increase in conscious recollection. Once again, two processes would cancel each other out. Of course, this interpretation is speculative, but not unlikely and it illustrates the general problem that occurs if several unobservable processes influence the same variable.

The hypothesised effects of the tendency to answer consistently on the hit rate can be retraced in Figure 6, which illustrates the relation between the conditional probabilities and the overall hit rate. The left column illustrates the response probabilities estimated from the observed data. The right column illustrates the hypothetical response probabilities that should occur without the strategy to answer consistently (hypothetical values in grey rectangles). Assuming that some acceptances of a category question are based on recollection of the corresponding target, the hit rate was hypothesised to be higher after acceptance of the preceding category question compared to a rejection (compare Figure 6, .85 after accepted category question vs. .80 after rejected category question).

The hypothetical values were chosen to approximate the observed data but nevertheless to illustrate that the observed data reflect the influence of a strategy to answer consistently whereas the hypothetical probabilities do not. Thus, the hypothetical probability of accepting a target after an accepted category question had to be lower compared to the corresponding observed hit rate (.85 vs .89). Accordingly, the hypothetical probability of rejecting a target after a rejected category question had to be lower compared to the corresponding observed false rejection rate (.20 vs .42). Consistent with predictions, the observed overall hit rate (.82) was lower compared to the hypothetical hit rate (.84). Thus, the example illustrates that the contribution of the strategy to answer consistently can decrease the hit rate for cued targets.

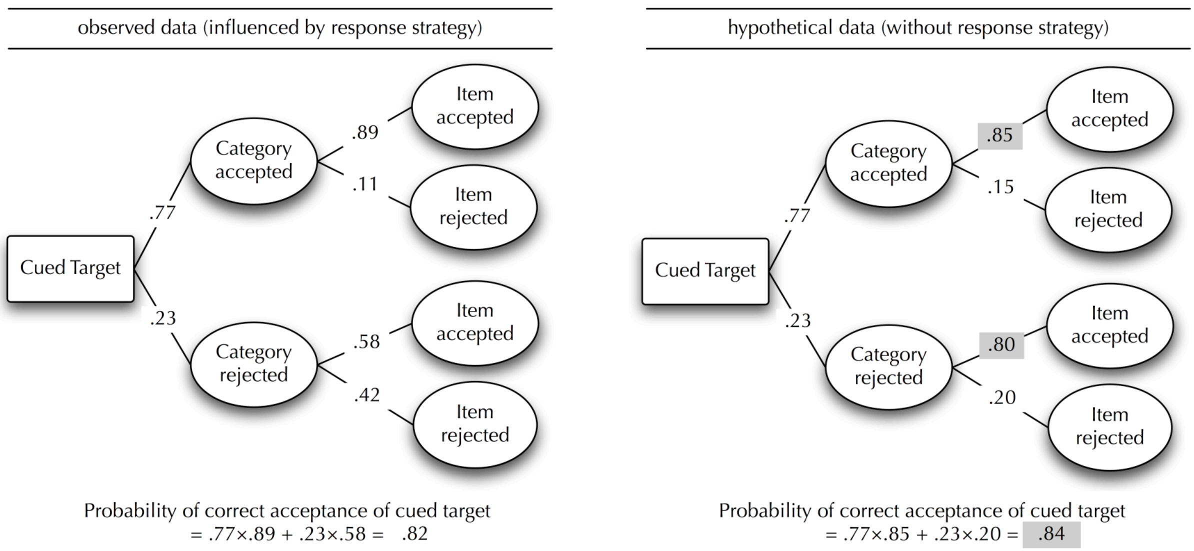

Figure 6 Illustrative example for the influence of the strategy to answer in a manner consistent with the preceding category question on the overall hit rate. The figure presents the possible answers to cued targets and their preceding category questions. It is assumed that the relative frequencies depicted on the left hand side are influenced by the response strategy whereas the hypothetical probabilities depicted on the right hand side are not. The example illustrates that the response strategy can result in a decrease of the hit rate.

To summarise, the results of Experiment 1 and 2 show that cueing items by the corresponding category name can not only reduce but even eliminate the false recognition effect. This indicates that category cueing was successful in increasing the probability of the recall-to-reject process by facilitating conscious recollection of learned items. However, the increased level of false alarms for cued items and the results of the conditional analysis lead to the following assumption: The cueing procedure used in Experiment 1 and 2 gave rise to side effects that work in opposition to recall-to-reject, namely a simultaneous increase of familiarity and a tendency to answer consistently. However, as recall-to-reject as well as its side effects affect recognition data, disentangling the processes seems to be virtually impossible without using additional methods like multinomial modeling. The author believes that Experiment 1 and 2 still provide clear evidence for recall-to-reject (this will be discussed in more detail in the General Discussion). However, it would be worthwhile to find a cueing manipulation where the effect of recall-to-reject is not impaired by side effects.

9 Experiment 3

The main goal of Experiment 3 was to eliminate the side effects of category cueing in order to achieve an absolute reduction of the false alarm rate. This should be possible by simply using phonological association instead of semantic association as the similarity manipulation. Thus, associated distractors in Experiment 3 were words that were not presented during the study phase (e.g., house) but rhymed with a word of the study list (e.g., mouse). Since the combination of semantic cues and phonological association prevents a direct relation between category question (e.g., animal?) and the following item (e.g., house), the category question should not lead to an increase of familiarity for the following items. Furthermore, changing the similarity manipulation should also eliminate the tendency to answer consistent with the preceding category question. The relation between category question and the following item, which is now indirect and phonological, only becomes obvious if the learned category exemplar to which the category question refers is correctly recalled. For example, if the word *mouse* is learned and *house* is used as the associated distractor, the relation between the category question *animal?* and the following item *house* only becomes apparent if the word *mouse* is recalled. And it is exactly this situation where participants were expected to effectively use the recall-to-reject process. Therefore, the false alarm rate for cued associated distractors was expected to be lower than for non-cued associated distractors.

9.1 Pilot study

A pilot study was run to evaluate the item material that was created for Experiment 3. A simple old/new recognition test investigated if the phonologically associated item material fulfils the precondition to induce a false recognition effect (which could be reduced by category cueing later on). This cannot be taken for granted as some studies failed to observe a false recognition effect when rhymes were used as associated distractors (e.g., Brainerd et al., 1998).

9.1.1 Method

For the sake of comparability, the pilot study corresponded as far as possible to Experiment 3. Besides the item material that should be evaluated, the length of study and test lists and presentation parameters were the same for both experiments.

9.1.1.1 Participants

Participants were 24 students, 19 of whom were female. The mean age of participants was 20.92 (*SD* = 2.30, ranging from 18 to 27 years). All participants were native German speakers and were paid € 3.00 for their participation.

9.1.1.2 Materials

The item material consisted of 140 single category names with one category exemplar each (termed single category item pool) and 40 category pairs, that is, two categories whose exemplars rhymed (i.e., were phonologically similar) and were orthographically similar[11] (e.g., building–house; animal–mouse; termed category pairs item pool). Selection of stimuli was further constrained by the need to avoid that other category exemplars rhymed and that category exemplars belonged to more than one category. Furthermore the item set did not contain any compound words or homonyms. All category exemplars were German nouns with one to three syllables. See Appendix B for a complete list of the item material. Targets and standard distractors were taken from the single category item pool. Associated distractors were taken from the category pairs item pool.

Participants studied five lists, each of which consisted of a total of 30 items. All words in the study list were presented in random order, except for the constraint that the first and last five items were primacy and recency buffers. Each of the five test lists consisted of 40 items presented in random order. Test lists comprised twelve targets, sixteen standard distractors and eight associated distractors. Additionally, in order to adjust the percentage of target items in the test (which was 40% in Experiment 3), two out of the five primacy and recency buffers were tested but not included in the analysis.

---

[11] Although the linguistic definition of rhymes is only based on articulation and not on typeface, rhymes that are pronounced similarly but written differently (e.g., key, sea) were excluded in order to maximise orthographical similarity. Additionally, the number of syllables was the same for both items of a rhyme pair. Hence, only the consonants of the initial sound differed between items of a rhyme pair.

9.1.1.3 Procedure

The procedure as well as presentation parameters were identical to Experiment 3 (see below) with the exception that the recognition test was a simple old/new recognition test without any category questions.

9.1.1.4 Design

The pilot study employed a one-factorial design with item type (target, distractor, associated distractor) as within-subject variable. The relative frequency of an old response in the old/new recognition test was the dependent variable.

The comparison between the false alarm rate for associated distractors and for standard distractors defined the false recognition effect. Given $N = 24$ participants and $\alpha = .05$, it was possible to detect a false recognition effect of medium size $d_z = 0.5$ with a probability of $1 - \beta = .78$.

9.1.2 Results

The false alarm rate for associated distractors ($M = .137$, $SE = .015$) was significantly higher than for standard distractors ($M = .079$, $SE = .012$), that is, a typical false recognition effect was observed ($t(23) = 4.53$, $p < .001$, $d_z = 0.94$). The hit rate for targets was .872 ($SE = .016$).

The finding of a phonological false recognition effect of large size (according to effect size conventions of Cohen, 1988) proves the applicability of the item material created for Experiment 3. Thus, it can be investigated in Experiment 3 if the increased false alarm rate found for phonologically associated items can be reduced by category cueing.

9.2 Method

9.2.1 Participants

Participants were 64 adults (mostly students), 46 of whom were female. The mean age of participants was 23.59 ($SD = 4.43$), ranging from 19 to 38 years. All participants were native German speakers and none had participated in Experiment 1 or 2 or in the pilot study. Participants were tested in groups of up to four and were paid € 3.00 for their participation.

9.2.2 Materials

The same item material as in the pilot study was used (compare Appendix B). As in the previous experiments, participants studied category exemplars and were later tested for recognition memory. Examples of all item types are shown in Table 5. Test items were targets, standard distractors and associated distractors. As in Experiment 1, all items were preceded by a category question where participants had to decide whether an exemplar of the presented category had been in the study list or not. Similar to Experiment 1, the category questions preceding all non-cued items were unrelated to the following test items[12]. The category question for cued associated distractors referred to the studied exemplar that rhymed with the following test item. There was no cueing condition for standard distractors or targets. Standard distractors cannot be cued indirectly as there is no studied item that could be cued. In order to create an indirect cueing condition for targets one would have to present rhyme pairs in the study list, which the author decided against, since this should reduce the effectiveness of the recall-to-reject process (see Chapter 5.5.2 of the Introduction as well as Chapter 12.3 of the General Discussion for the role of mutual exclusivity of stimuli). Please note that in contrast to Experiment 1 and 2, a test item was never an exemplar of the preceding category.

[12] As in Experiment 1, it was additionally varied whether an item of the unrelated category under question was learned or not.

Table 5

Examples for the different item conditions in Experiment 3. For simplicity, the table does not show items of the study list referring to the category question of non-cued items.

Item condition	Study list	Test list	
	Item referring to item question	Category question	Item question
Non-cued targets	house	sports?	house?
Non-cued distractors	–	sports?	house?
Non-cued associated distractors	mouse	sports?	house?
Cued associated distractors	mouse	animal?	house?

Targets and standard distractors were taken from the single category item pool. Cued as well as non-cued associated distractors were taken from the category pairs item pool. Stimulus material for the unrelated category question prior to non-cued targets, standard and associated distractors was also taken from the single category item pool. Categories were randomly allocated to item types and category questions for each subject. For associated distractors, it was decided at random which exemplar was presented in the study list and which was presented as the associated distractor in the test list.

The author is aware that the use of separate item pools can basically confound comparisons. However, the use of two separate item pools seemed justified for several reasons. Firstly, all associated distractors, and therefore all items for the critical comparison (which is between cued and non-cued associated distractors) are taken from the same item pool. Secondly, item pools were comparable with respect to the number of syllables and frequency[13]. Thirdly, the quantity of category pairs is very limited and thus the need for a separate item pool for less important items is inevitable to assure an adequate number of trials for analysis to be based on.

---

[13] The number of syllables ranged from 1 to 3 (M = 2.07, SD = 0.62) for the single category item pool vs. 1 to 2 (M = 1.50, SD = 0.50) for the category pairs item pool. Mean word frequency (according to Wortschatz Lexikon Deutsch, 2005) was M = 12.54 (SD = 2.27) for the single category item pool vs. M = 12.19 (SD = 2.38) for the category pairs item pool.

Participants studied five lists, each of which consisted of a total of 30 items. All words in the study list were presented in random order, except for the first and last three items, which were primacy and recency buffers.

Each of the five test lists consisted of 20 items preceded by their respective category question. All items of the test list were presented in random order. Each test list was composed of eight non-cued targets, four non-cued distractors, four non-cued associated distractors, and four cued associated distractors.

9.2.3 Procedure

The procedure was identical to that of Experiment 1 with the exception that words in the study phase were presented for 600 ms each, with a blank screen shown for 400 ms between two word presentations. Like in Experiment 1, participants were not informed about details of study and test list construction (that words which rhyme do not occur during study and that, as in Experiment 1, only one exemplar of each category was presented during study).

9.2.4 Design

Experiment 3 employed a one-factorial design with item type (non-cued target, non-cued distractor, non-cued associated distractor, cued associated distractor) as within-subject variable. The relative frequency of an old response in the old/new recognition test for the different item conditions was the dependent variable.

The comparison between the false alarm rate for associated distractors and for standard distractors defined the false recognition effect. Given $N = 64$ participants and $\alpha = .05$, it was possible to detect a false recognition effect of large size $d_z = 0.94$ (according to the pilot study) with a probability of $1 - \beta > .99$. In contrast to Experiments 1 and 2, the false recognition effect could only be computed for non-cued items since the design did not include cued standard distractors. However, the comparison between the false alarm rate for non-cued and cued associated distractors defined the effect of cueing on recall-to-reject. Please note that this comparison is rather conservative as recall-to-reject is expected to reduce the false alarm rate for associated distractors in the present Experiment but an increase was observed in Experiments 1 and 2 due to side effects.

9.3 Results

Participants answered 67% of the category questions (*SE* = .010) and 85% of the item questions (*SE* = .010) correctly. The relative frequencies of old responses for the different item types are shown in Table 6[14].

Table 6

Proportion of hits and false alarms for different item conditions in Experiment 3 (standard errors in parentheses). The table shows the relative frequencies of old responses both in overall terms and depending on the answer to the preceding category question. For the conditional analysis, *t*-values, degrees of freedom (*df*) and effect sizes (d_z) are also listed.

Item condition	Unconditional	Conditional			
		Preceding category accepted	Preceding category rejected		
	M (*SE*)	*M* (*SE*)	*M* (*SE*)	*t* (*df*)	d_z
Non-cued targets	.813 (.014)	.795 (.018)	.831 (.016)	1.81 (63)	0.23
Non-cued distractors	.107 (.013)	.115 (.016)	.099 (.016)	0.89 (63)	0.11
Non-cued associated distractors	.153 (.017)	.158 (.019)	.162 (.022)	0.26 (63)	0.03
Cued associated distractors	.125 (.016)	.098 (.016)	.187 (.026)	4.04 (63) *	0.51

* $p < .001$

The false alarm rate for non-cued associated distractors was significantly higher than for non-cued standard distractors, that is, a typical false recognition effect was observed ($t(63) = 3.10$, $p = .001$, $d_z = 0.39$). The false alarm rate for cued associated distractors was significantly lower than that for non-cued associated distractors ($t(63) = 2.14$, $p = .018$, $d_z = 0.30$). A comparison between cued associated distractors and non-cued standard distractors revealed no differences in the false alarm rates indicating that category cueing eliminated the false recognition effect ($t(63) = 1.42$, $p = .159$, $d_z = 0.18$).

[14] Preliminary analysis confirmed that the relative frequencies of old responses for non-cued targets, distractors, and associated distractors did not differ depending on whether an exemplar of the preceding category was learned or not ($t(63) = 0.23$, $p = .819$, $d_z = 0.03$ for targets, $t(63) = 1.83$, $p = .072$, $d_z = 0.23$

Again, recognition data were additionally analysed depending on the answer to the preceding category question. Predictions for non-cued items are the same as in Experiment 1 and 2: As preceding category questions for non-cued items are unrelated to the following item, acceptance or rejection of the preceding category question is not expected to have any effect on the following item recognition decision. For cued associated distractors the prediction of Experiment 1 is resumed. The false alarm rate is expected to be lower if the preceding category question had been accepted. This is expected to result from recall-to-reject because if the preceding category question has been accepted based on recall of the corresponding studied category exemplar, the discrepancy to the following associated distractor is likely to be detected and the associated distractor will be rejected. When successful in eliminating the side effects of category cueing, this time, the conditional results for cued associated distractors should conform to predictions.

The conditional analysis (see Table 6) revealed that the answer to the preceding category question had an effect on the probability of an old response for cued associated distractors only. The false alarm rate for cued associated distractors was significantly lower if the category question had been accepted.

9.4 Discussion

Replicating the results of Experiments 1 and 2, cueing associated distractors by a category question about the corresponding studied item eliminated the false recognition effect. Moreover, in Experiment 3, the false alarm rate for cued associated distractors was reduced significantly down to the level of that for non-cued standard distractors. This finding indicates that the side effects observed in Experiments 1 and 2 could be successfully eliminated by using phonological association. Furthermore, the results support the interpretation of Experiment 1 and 2 in that category cueing increases the probability of conscious recollection and thereby facilitates the recall-to-reject process.

---

for distractors, and $t(63) = 0.86$, $p = .395$, $d_z = 0.11$ for associated distractors). All subsequent analyses on non-cued items could therefore be performed on collapsed data.

These conclusions are further confirmed by the results of the conditional analysis. The false alarm rate for cued associated distractors was lower when the preceding category question had been accepted compared to when it had been rejected. This was expected since the recollection of a category exemplar should lead to the acceptance of the corresponding category question and make recall-to-reject and, hence, the rejection of the following associated distractor very likely.

To summarise, by eliminating possible side effects of category cueing (increase of familiarity and tendency to answer consistently) the elimination of the false recognition effect found in Experiments 1 and 2 could be replicated and moreover an absolute reduction of the false alarm rate could be achieved without giving explicit recall-to-reject instructions.

The recall-to-reject process was very effective in the present experiments, as the false recognition effect was not only reduced but eliminated. However, the structure of the study list may have played an important role in this effectiveness. In the present experiments, associated distractors were similar to only one item of the study list. Therefore, the conscious recollection of the corresponding presented item could be taken as strong evidence that the associated distractor was not learned despite its high familiarity. If more associated items are presented during study (up to five in Gallo, 2004), the associated distractor can only be rejected with high confidence if participants know how many similar items were presented and if they exhaustively recall all presented items. Hence the effectiveness of the recall-to-reject process should decrease with an increasing number of similar items and/or a variable number of items per category (Brainerd et al., 2003; Gallo, 2004). Consequently, the usefulness of category cues in reducing the false recognition effect should decrease as well. As discussed in Chapter 5.5.2, mutual exclusiveness of stimuli has also been suggested to be a necessary precondition for recall-to-reject.

10 Experiment 4

In Experiment 4, mutual exclusivity of stimuli was excluded by including pairs of phonologically associated items (i.e., rhyme pairs) in the study list. To increase participants' awareness of the presentation of rhyme pairs during study, some rhyme pairs were also included in the test list. It was expected that the absolute reduction of the false alarm rate for cued compared to non-cued associated distractors that was observed in Experiment 3 would be diminished or absent in Experiment 4. Furthermore, concerning the conditional analysis for cued associated distractors, predictions differ from Experiment 3. As including rhyme pairs in study lists was expected to reduce the effectiveness of recall-to-reject, the lower false alarm rate after acceptance of the preceding category question that was observed in Experiment 3 and attributed to recall-to-reject, was expected to be absent.

10.1 Method

10.1.1 Participants

Participants were 63 adults (mostly students), 35 of whom were female. The mean age of participants was 22.46 (*SD* = 3.96), ranging from 18 to 43 years. All of them were native German speakers. None of them had participated in the preceding experiments. Participants were tested in groups of up to four and were not paid for participation since psychology students conducted participant recruitment as part of the requirements for obtaining course credit.

10.1.2 Materials

The item material of Experiment 3 was used and extended for the additional presentation and testing of rhyme pairs. The additional item material consisted of 20 rhyme pairs without corresponding category names (e.g., hair; pair) and 20 rhyme pairs with only one category name (e.g., criminal–thief; grief).

The test lists additionally comprised non-cued rhyme pairs (e.g., drug?–hair?; artist?–pair?). Half of the non-cued rhyme pairs were presented during study (i.e., rhyme pair targets) and the other half was not presented during study (i.e., rhyme pair distractors). All non-cued rhyme pairs were taken from the rhyme pairs item pool without corresponding category name. The test list also comprised cued rhyme pair targets, that is, a studied rhyme pair item which is cued by the category name corresponding to the rhyming item (e.g., if *thief* and *grief* were learned and *grief* was presented during test, the preceding category question asked whether a *criminal* was learned).

Cued rhyme pair targets were presented equally often as the first or second item of the studied rhyme pair. For non-cued rhyme pairs an exemplar of only one of the two categories was always studied. It was balanced if an exemplar of the first or second category was studied. For non-cued rhyme pairs the order of presentation of rhyme pair targets was the same in the study and test phase.

Participants studied five lists, each of which consisted of a total of 42 items. In contrast to Experiment 3, study lists additionally contained six rhyme pairs. Half of the rhyme pair targets were presented in immediate succession to attract participants' attention. The other half of rhyme pair targets were presented with two unrelated items interspersed. This was done to increase participants' uncertainty as to whether an item had been presented with or without a rhyme word. Either one or both rhyme pair targets were tested in the following recognition task.

All words in the study list were presented in random order, except for the constraint that half of the rhyme pairs were presented in immediate succession and the other half were presented with two unrelated items interspersed. As in Experiment 3, the first and last three items of the study list were primacy and recency buffers.

Each of the five test lists consisted of 28 items preceded by their respective category question. All items of the test list were presented in random order. Concerning the item categories of Experiment 3, tests lists of Experiment 4 comprised four non-cued targets, four non-cued distractors four non-cued associated distractors, and four cued associated distractors. Additionally, the test list of Experiment 4 included four cued rhyme pair targets, four non-cued rhyme pair targets (i.e., two learned rhyme pairs), and four non-cued rhyme pairs distractors (i.e., two unlearned rhyme pairs).

10.1.3 Procedure

The procedure was identical to that of Experiment 3 except for an increased presentation time. To compensate for the increased list length (42 items in Experiment 4 vs. 30 items in Experiment 3), study items in Experiment 4 were presented for 1000 ms each, with a blank screen shown for 300 ms between two word presentations. Participants were not informed about details of study list construction.

10.1.4 Design

The design of Experiment 4 was nearly identical to that of Experiment 3 with the exception that study and test lists in Experiment 4 included rhyme pairs. However, these additional item types (i.e., non-cued rhyme pair targets, non-cued rhyme pair distractors, and cued rhyme pair targets) were not incorporated in the design and recognition data are not presented in the present work[15].

The comparison between the false alarm rate for associated distractors and for standard distractors defined the false recognition effect in the non-cued condition. Given $N = 63$ participants and $\alpha = .05$, it was possible to detect a false recognition effect of size $d_z = 0.94$ (according to the pilot study of Experiment 3) with a probability of $1 - \beta > .99$. Additionally, the comparison between the false alarm rate for non-cued and cued associated distractors defined the effect of cueing on recall-to-reject. Given an effect of size $d_z = 0.30$ (according to Experiment 3), it was possible to detect an effect of cueing on recall-to-reject with a probability of $1 - \beta = .76$.

[15] Recognition data of rhyme pair items were not of relevance to the hypothesis. Furthermore, within rhyme pair item types, the variation of presenting rhyme pairs in immediate succession or with two unrelated items interspersed and the variation of testing either the first or second word of a rhyme pair would have to be taken into account. The number of trials ranges from only five to ten per participant, which is too little to obtain reliable results.

10.2 Results

Participants answered 65% of the category questions (*SE* = .010) and 78% of the item questions (*SE* = .010) correctly. The relative frequencies of old responses for the same item types as in Experiment 3 are shown in Table 7[16].

Table 7

Proportion of hits and false alarms in Experiment 4 for the item conditions of Experiment 3 (standard errors in parentheses). The table shows the relative frequencies of old responses both in overall terms and depending on the answer to the preceding category question. For the conditional analysis, *t*-values, degrees of freedom (*df*) and effect sizes (d_z) are also listed.

	Unconditional	Conditional			
Item condition		Preceding category accepted	Preceding category rejected		
	M (*SE*)	*M* (*SE*)	*M* (*SE*)	*t* (*df*)	d_z
Non-cued targets	.752 (.019)	.742 (.024)	.759 (.021)	0.99 (61)	0.13
Non-cued distractors	.094 (.011)	.106 (.015)	.111 (.018)	0.2 (61)	0.04
Non-cued associated distractors	.229 (.019)	.240 (.025)	.209 (.023)	1.07 (61)	0.14
Cued associated distractors	.227 (.019)	.226 (.025)	.255 (.027)	0.70 (60)	0.09

* $p < .001$

The false alarm rate for non-cued associated distractors was significantly higher than for non-cued standard distractors, that is, a typical false recognition effect was observed ($t(62) = 8.14$, $p < .001$, $d_z = 1.03$). There was no difference between the false alarm rate for cued associated distractors and non-cued associated distractors ($t(62) = 0.09$, $p = .932$, $d_z = 0.01$). A comparison between cued associated distractors and non-cued standard distractors revealed that the false alarm rate for cued associated distractors was significantly higher than for non-cued standard distractors ($t(62) = 7.44$, $p < .001$, $d_z = 0.94$), that is, a false recognition effect was observed despite category cueing.

[16] Preliminary analysis confirmed that the relative frequencies of old responses for non-cued targets, distractors and associated distractors did not differ depending on whether an exemplar of the preceding category was learned or not ($t(62) = 1.66$, $p = .102$, $d_z = 0.21$ for targets, $t(62) = 0.73$, $p = .470$, $d_z = 0.09$

With respect to the conditional analysis, results for non-cued items were identical to all three previous experiments and conformed to predictions: Acceptance or rejection of the preceding category question did not have any effect on the following item recognition decision. As expected, the results for cued associated distractors differed from Experiment 3. Table 7 shows that, contrary to Experiment 3, the false alarm rate for cued associated distractors did not differ depending on the answer to the preceding category question.

10.3 Discussion

When rhyme pairs were included in study lists, category cueing did not reduce the false recognition effect. Rhyme pairs were additionally included in the test lists to increase participants' awareness of the presentation of rhyme pairs during study. This demonstrates that the structure of study lists has an important influence on the occurence of recall-to-reject. In the present experiment, category cueing might still have facilitated conscious recall of the corresponding presented item. However, this could not be taken as evidence that the associated distractor was not learned because study lists did include pairs of associated items. Therefore, the structure of the study list did prevent the usefulness of category cueing with respect to recall-to-reject.

To summarise, Experiment 4 demonstrates the importance of study list structure with respect to mutual exclusivity of stimuli for the usefulness of recall-to-reject. If the structure of the study list prevents conscious recall of associated learned items from providing compelling evidence of detecting associated distractors, then category cues cannot be used to increase recall-to-reject and therewith reduce the false recognition effect.

---

for distractors, and $t(62) = 0.70$, $p = .488$, $d_z = 0.09$ for associated distractors). All subsequent analyses on non-cued items could therefore be performed on collapsed data.

11 Analyses of Response Latencies

In addition to recognition data, the time needed for the old/new decision for item questions was analysed. As the recognition task was self-paced, responses times have to be interpreted with caution. But an analysis seemed reasonable as it was ensured that participants responded with their index fingers and that these remained on the keyboard during the duration of the experiment.

As variability among participants mean response latencies was quite high, a cutoff value that is based on individual participants' means and standard deviations was chosen (as recommended by Ratcliff, 1993). Response latencies shorter than 200 ms or longer than two standard deviations above participants mean were eliminated as outliers. The correction for response latency outliers resulted in an elimination of a reasonable amount of data (4.97% (*SD* = 1.67) in Experiment 1, 4.97% (*SD* = 5.19) in Experiment 2, 4.98% (*SD* = 1.55) in Experiment 3, and 5.03% (*SD* = 1.32) in Experiment 4). This supports the choice of cutoff criterion since a reasonable range is to eliminate 5% to 10% of the data (Ratcliff, 1993). On average, participants mean response latency for item questions was 1439 ms (*SD* = 362) in Experiment 1, 1363 ms (*SD* = 291) in Experiment 2, 1611 ms (*SD* = 497) in Experiment 3, and 1439 ms (*SD* = 408) in Experiment 4.

If category cueing provides information relevant to the following item recognition decision, overall response latencies for cued items should be reduced compared to uncued items. This prediction is in line with the finding that recognition judgments are facilitated when primed with an episodic task (Lewandowsky, 1986). Furthermore, when items are category exemplars of the preceding category question (Experiments 1 and 2), the prediction of shorter response latencies for cued items is in line with expectancy-based priming effects (for an overview, see Neely, 1991). More specifically, for associated distractors, shortest response latencies should be observed when recall-to-reject does occur. Without cueing manipulation, longer response times would be indicative of recall-to-reject (Gallo, 2003). In the present study, however, as a result of cueing, the time-consuming recollection process takes place before the test item is presented for recognition.

Thus, if category cueing does increase the probability of recall-to-reject, as supposed in Experiment 1 to 3, then the average response latency for cued associated distractors is hypothesised to be shorter than for non-cued associated distractors. In Experiment 4, where including rhyme pairs in the study list should have prevented category cueing from facilitating recall-to-reject, the average response latency for cued associated distractors is not expected to be shorter than that for non-cued associated distractors.

Similar to recognition data, additional hypothesis can be established for response latencies analysed depending on the answer to the preceding category question (conditional analysis). As the probability of recall-to-reject is supposed to be higher if the category question preceding cued associated distractors is accepted, the average response latency for cued associated distractors is expected to be shorter after accepted compared to rejected category question in Experiments 1 to 3. In Experiment 4 where mutual exclusivity of stimuli was excluded, no such difference in response latencies is expected. As the category questions for non-cued associated distractors were unrelated to the following items in Experiments 1, 3, and 4 (non-cued associated distractors were not preceded by any category question in Experiment 2), no such difference is expected for non-cued associated distractors.

11.1 Results

Figure 7 shows response latencies for the different item conditions in Experiment 1 to 4. The responses to non-cued targets were significantly faster than to non-cued standard distractors, a result typically found in recognition tasks (e.g., Lewandowsky, 1986; Taylor & Juola, 1974); $t(73) = 4.19$, $p < .001$, $d_z = 0.98$ in Experiment 1, $t(74) = 7.17$, $p < .001$, $d_z = 1.67$ in Experiment 2, $t(63) = 10.42$, $p < .001$, $d_z = 2.63$ in Experiment 3, $t(62) = 6.96$, $p < .001$, $d_z = 1.77$ in Experiment 4.

At a descriptive level, in Experiment 1 and 2, response latencies for cued items were shorter than for non-cued items. Paired *t*-Tests confirmed that average response latency for cued associated distractors was shorter than for non-cued associated distractors ($t(73) = 9.96$, $p < .001$, $d_z = 1.17$ in Experiment 1 vs. $t(74) = 13.37$, $p < .001$, $d_z = 1.55$ in Experiment 2). As expected, this also applied to Experiment 3 ($t(63) = 2.21$, $p = .016$, $d_z = 0.28$) but not to Experiment 4 ($t(62) = 1.22$, $p = .227$, $d_z = 0.16$).

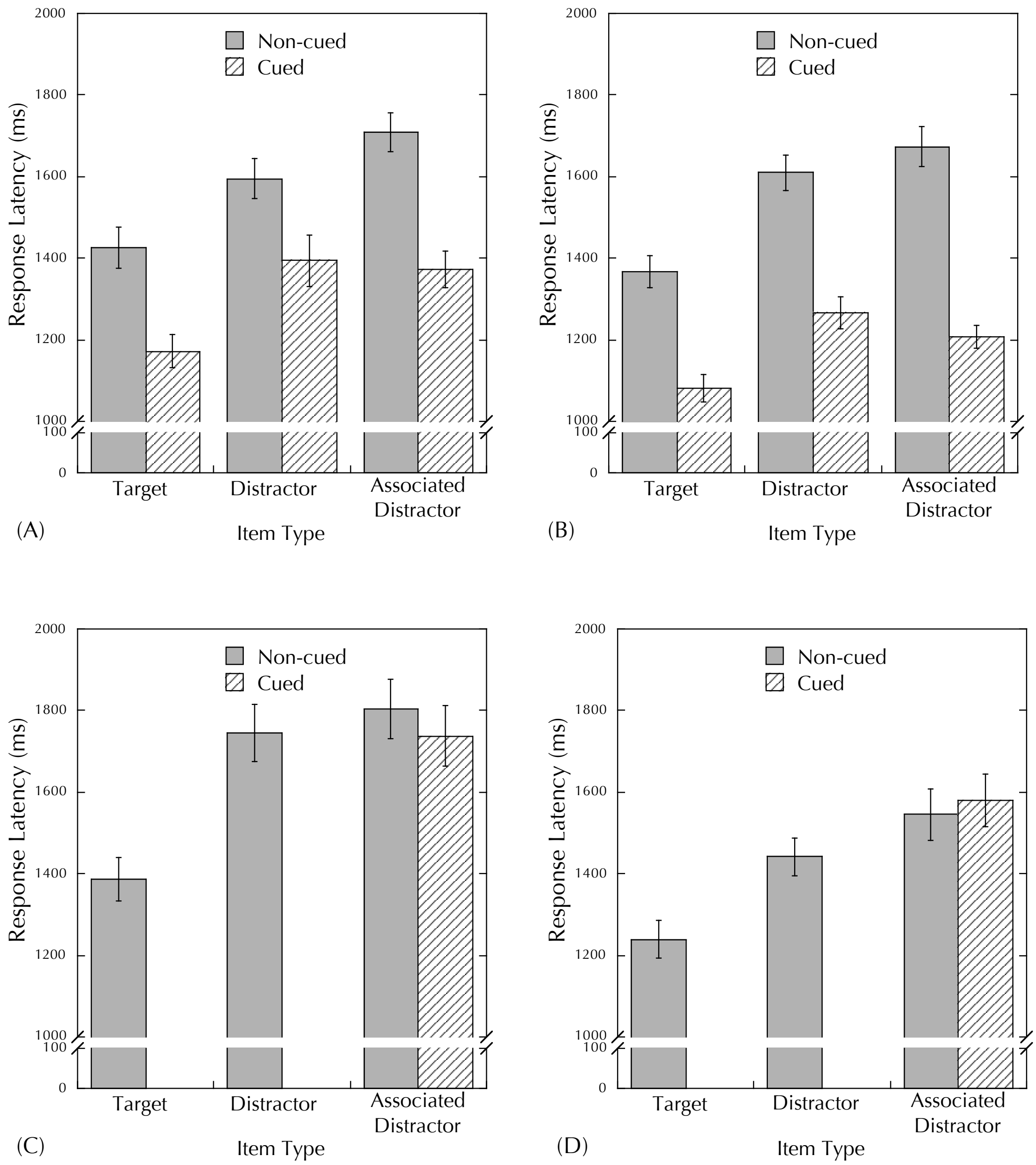

Figure 7 Response latencies to item questions of the different item conditions in (A) Experiment 1, (B) Experiment 2, (C) Experiment 3, and (D) Experiment 4 (error bars indicate standard errors of the mean).

Response latencies were additionally analysed depending on the answer to the preceding category question. In Experiments 1, 3, and 4, as expected, the answer to the category question preceding non-cued associated distractors did not have any effect on the response latency for the following recognition decision ($t(72) = 1.52$, $p = .134$, $d_z = 0.18$ in Experiment 1, $t(63) = 1.10$, $p = .275$, $d_z = 0.14$ in Experiment 3, $t(61) = 0.54$, $p = .591$, $d_z = 0.07$ in Experiment 4). Consistent with recall-to-reject predictions, the response latency for cued associated distractors was shorter if the preceding category question had been accepted compared to if it had been rejected

($t(62) = 2.77$, $p = .004$, $d_z = 0.35$ in Experiment 1, $t(68) = 3.24$, $p = .001$, $d_z = 0.39$ in Experiment 2, $t(62) = 6.51$, $p < .001$, $d_z = 0.83$ in Experiment 3). Contrary to expectations, the same difference for cued associated distractors was found in Experiment 4 ($t(60) = 2.68$, $p = .010$, $d_z = 0.35$).

11.2 Discussion

The overall response latency for cued associated distractors was shorter than that for non-cued associated distractors in Experiments 1 to 3, but not in Experiment 4. This pattern of results is consistent with the interpretation of recognition data, that is, category cueing increased the probability of recall-to-reject in Experiments 1 to 3, but not in Experiment 4.

However, comparing the effect size measure d_z, the reduction of response latencies for cued compared to non-cued associated distactors was strikingly larger in Experiments 1 and 2 compared to Experiment 3 ($d_z = 1.17$ and $d_z = 1.55$ in Experiment 1 and 2, respectively vs. $d_z = 0.28$ in Experiment 3). This could be due to the tendency to answer consistently which is supposed to be elicited as a side effect of category cueing. It seems plausible that responses based on simple response strategies such as response repetition or alternation should be very fast because responses should depend on the previous response rather than the item. Therefore, response latencies influenced by response strategies should be shorter than those based on memory processes alone. One could argue that the response latency difference between cued and non-cued associated distractors might solely be due to the tendency to answer consistently in Experiment 1 and 2. However, d_z values indicate that the effect was even larger in Experiment 2 compared to Experiment 1. This is consistent with recognition data that indicated that the probability of recall-to-reject was further increased by the modified cueing procedure of Experiment 2. Consequently, the reduction of response latencies for cued compared to non-cued associated distractors seems to be at least partly driven by recall-to-reject.

The interpretation of the unconditional response latency data in favour of recall-to-reject is further supported by the conditional analysis depending on the answer to the preceding category questions in Experiments 1 to 3. One might argue that the unexpected difference found for cued associated distractors in Experiment 4, where recall-to-reject was not expected, weakens the interpretation of conditional results in Experiments 1 to 3. In this matter, it would be interesting to perform an analysis of response latencies depending on the recognition decision as well as depending on the answer to the preceding category questions. Predictions about response latencies can be specified with respect to a *double conditional analysis*. As shortest response latencies are expected when recall-to-reject does occur, average response latency for cued associated distractors should be shortest for rejected items if the preceding category question has been accepted. This pattern of results should be found for Experiment 1 to 3 but not for Experiment 4. Unfortunately, an analysis of response latencies depending on the recognition decision as well as the answer to the preceding category questions would be based on too few trials to obtain reliable results[17].

When interpreting response latency differences, response sequences have to be taken into account. However, the shorter response latency for associated distractors found after accepted compared to rejected category questions cannot be attributed to a possible response repetition facilitation because the majority of cued associated distractors is rejected (the proportion varies between 0.856 and .773 for Experiments 1 to 4). Overall, the results of response latency analysis are consistent with recognition data and therefore provide converging evidence for recall-to-reject.

[17] For example, in Experiment 3, the number of false alarms for cued associated distractors for which the preceding category question had been accepted ranged from only 0 to 6 per participant (1.23 on average).

12 General Discussion

The present work focused on recall-to-reject, one of the central memory editing mechanisms predicted to prevent the occurrence of false memories. Recall-to-reject occurs when the correct rejection of an associated distractor is based on the recall of the corresponding studied item. As described in detail in the Introduction, recall-to-reject has been studied using a variety of different tasks and methods and multiple manipulations have been used to influence the probability of the recall-to-reject process. As discussed, previous research lacks evidence for recall-to-reject in item recognition as well as evidence based on recognition data whereas evidence from multinomial modeling and ROC analysis is numerous. The present work aimed at filling this gap and providing evidence for the hypothesis that facilitating recall-to-reject should result in a reduction of the false recognition effect in an item recognition task. A category cueing manipulation that has been shown to improve recall performance was used in all experiments to increase the probability of recall-to-reject.

Experiment 1 tested the hypothesis that category cueing can increase recall-to-reject and consequently reduce the false recognition effect even in an item recognition task and without explicit instructions. The results confirmed this prediction but indicated that category cueing simultaneously increases familiarity and induces a tendency to answer consistently. Both side effects will be discussed in the following Chapters, as will implications for previously published studies. Experiment 2 aimed at reducing the familiarity increase and maximise the effectiveness of recall-to-reject by providing participants with explicit instructions. However, Experiment 2 failed to reduce the familiarity increase but results suggest that explicit recall-to-reject instructions might have increased the base rate of recall-to-reject. The implications of the present experiments with respect to the importance of instructions for recall-to-reject will be discussed in the following. In Experiment 3 an item recognition task using phonological association was developed to eliminate the side effects of category cueing observed in Experiments 1 and 2. Results indicated that the design successfully prevented the occurrence of side effects and replicated that category cueing can be used to increase recall-to-reject and eliminate the false recognition effect even without explicit instructions. Finally, when mutual exclusivity of stimuli was removed in Experiment 4, category cueing no longer had an effect on false alarms to associated distractors. The importance of mutual exclusivity of stimuli for the utility of recall-to-reject will also be discussed in the following. Furthermore, methodological, theoretical and practical implications of the present work will be discussed. Finally, the General Discussion will close with a summary of the conclusions that can be drawn from the results of the present work.

12.1 Side Effects of Recall-to-reject Manipulations

The logic of most manipulations, like the category cueing manipulation used in the present work, is that the probability of recall-to-reject is tied to the probability of conscious recollection. Enhancing the probability of conscious recollection should increase the probability of recall-to-reject and consequently reduce the false recognition effect. The results of the present work confirmed this prediction but demonstrated that side-effects, which oppose the reduction of false alarms to associated distractors, can occur.

12.1.1 Familiarity Increase

The high level of false alarms to cued standard distractors that was observed in the first two experiments was indicative of an increase in familiarity. It was assumed that the high familiarity is due to the categorical overlap between category and item question and should apply to all types of cued items. This assumption is consistent with the finding that the false alarm rate for critical lures in the DRM paradigm increases with the number of related items that are tested before the critical lure (Coane & McBride, 2006). The finding that the same effect is present for unstudied lists indicates that activation during test contributes to the familiarity and subsequent false recognition. Marsh and Dolan (2007) provided further evidence of test induced priming effects (but see also Dodd, Sheard, & MacLeod, 2006; Marsh, McDermott, & Roediger, 2004), although a higher false alarm rate after the testing of related items was only observed in a speeded test condition. When the test was self-paced, an effect of previous testing of related items was only observed with respect to response times (see also Lewandowsky, 1986; Taylor & Juola, 1974).

A test induced priming effect might be especially likely in the first two experiments, where the testing of a category exemplar was preceded by the corresponding category question, which required participants to indicate if an exemplar of the category had been studied or not. Deciding if an exemplar from a specific category has been studied can be seen as a special case of categorical frequency estimation task (where the number of exemplars from a particular category ranges only from zero to one). As categorical frequency estimations seem not to be based on information stored at the superordinate level but on the retrieval of individual memory traces (Greene, 1989), it seems likely that memory traces of different category exemplars are activated during the category question. This could occur through a process similar to a generate-recognize strategy (Anderson & Bower, 1972; Bahrick, 1970). Thus, false recognition of the following item can be regarded as a kind of source monitoring failure because false recognition could be avoided if the increased familiarity of category exemplar would correctly be attributed to the preceding category question. The increased familiarity of cued items could also be due to spreading activation (Collins & Loftus, 1975) or a better accessibility of the gist trace in the context of the fuzzy trace theory (e.g., Brainerd & Reyna, 2002a). Furthermore, it might be possible that the category exemplar and the corresponding category name are combined into a compound cue (Ratcliff & McKoon, 1988).

Irrespective of the process by which familiarity is increased, one aim of Experiment 2 was to facilitate monitoring the source of familiarity by replacing the unrelated control condition of Experiment 1 by an item recognition task without any preceding category question. The results were comparable with those of Experiment 1 and therefore indicate that the manipulation was not effective in facilitating monitoring the source of familiarity. However, the results of Experiment 3 suggest that a familiarity increase caused by the preceding category question can be prevented by using phonological association. In fact, an indirect familiarity increase cannot be excluded as no cued standard distractors could be included in the design. The corresponding false alarm rate would have been indicative of a familiarity increase in comparison to the false alarm rate of non-cued standard distractors. Nevertheless, as the false alarm rate for cued associated distractors was found to be lower than for non-cued associated distractors, the increase of recall-to-reject must have been stronger than a possible increase of familiarity.

Despite the high level of false alarms for cued distractors caused by the increase of familiarity found in Experiment 1 and 2, the elimination of the false recognition effect was nevertheless indicative of the occurrence of recall-to-reject. However, if no cued standard distractors were incorporated in the design, a null effect would have been observed, because the absolute level of the false alarm rate was the same for non-cued and cued associated distractors.

The simultaneous familiarity increase that was observed in the present work can explain why researchers often failed to find evidence for recall-to-reject in recognition data. Many manipulations that have been used to increase recall should lead to a parallel increase in familiarity. As familiarity and recall-to-reject have opposing effects on false recognition, both processes can cancel each other out. Behavioural data will only be indicative of recall-to-reject if its effect is stronger than the effect of familiarity or if data are compared to a control condition that is solely influenced by familiarity.

For example, a null effect of repeating items during study has been repeatedly observed with respect to behavioural data (Brainerd et al., 2006; T. C. Jones & Jacoby, 2001; Lampinen et al., 2004; Tussing & Greene, 1999). Tussing and Greene assumed that an increase of recall-to-reject was offset by a simultaneous increase of familiarity. This interpretation is in line with repetition priming effects (e.g., Neely, 1991) and is confirmed by the results of Jones and Jacoby, who found a null effect of repetition in a long deadline condition but an increase in false recognition in a short deadline condition. As recognition memory is assumed to be solely influenced by familiarity in a short response deadline condition, the observed increase in false recognition confirms the assumption that repetition simultaneously increases familiarity. Thus, the increased familiarity was opposed by recall-to-reject in the long deadline condition, resulting in a null effect that can be interpreted in favour of recall-to-reject.[18]

[18] Furthermore, the recall-to-reject interpretation of behavioural repetition data is in line with the finding that repetition increased the recall-to-reject parameter of ROC analysis (Lampinen et al., 2004) as well as the recall-to-reject parameter of the conjoint recognition model (Brainerd et al., 2006).

With respect to the cueing condition of Experiments 1 and 2, the increase of familiarity could be demonstrated by manipulations which should decrease the probability of recall-to-reject. For example, a higher false alarm rate for cued associated distractors compared to cued standard distractors should be observed in a short response deadline condition, when attention is divided during study or when pairs of category exemplars are included in the study list. The comparison of cueing conditions with and without the contribution of recall-to-reject would make the present results more compelling.

12.1.2 Tendency to Answer Consistently

A second side effect, which was observed in Experiments 1 and 2, was that participants tended to respond such that their answer to the item question was consistent with their answer to the preceding category question. As was obvious from the conditional analysis, the probability of falsely accepting a cued standard distractor was higher when the preceding category question had been accepted compared to when it had been rejected. For example, if participants had not studied any exemplar of the category *vegetables*, the probability of falsely accepting the item *spinach* was higher when the question whether an exemplar of the category *vegetables* had been learned or not had been answered with yes. As no exemplar of the category *vegetables* had been presented during study, the incorrect yes response to the category question as well as to the item question is likely to be due to guessing. The tendency to answer consistently is important for the interpretation of the results of Experiment 1 and 2 with respect to recall-to-reject. As discussed and illustrated by the hypothetical example in the Discussion of Experiment 1, the tendency to answer consistently should increase the false alarm rate for associated distractors but reduce the false alarm rate for standard distractors. Thus, the tendency to answer consistently should increase the false recognition effect and therefore oppose the recall-to-reject process.

Furthermore, the response tendency can explain the lack of an increase in the hit rate for cued compared to non-cued targets in Experiment 2. Based on the assumption that category cues increase the probability of conscious recall, the hit rate of targets was expected to be higher for cued compared to non-cued targets. However, as illustrated in the Discussion of Experiment 2, the tendency to answer consistently can lower the hit rate for cued targets. Thus, it seems likely that both effects opposed each other, resulting in a null effect.

In contrast to the familiarity increase, the tendency to answer consistently is a side effect that should be restricted to the specific design used in the present work. The utilisation of this response tendency is likely to be due to the fact that a response for the category question was required. This was done to maximise participants' effort and guarantee that participants were attending to the retrieval cues. It might be possible to prevent the occurrence of this tendency by including *don't know* as a response option or by not requiring participants to answer to the category question. Although this would not prevent a familiarity increase due to the categorical overlap between category and item question, it might reduce the level of false alarms for associated distractors.

12.2 Instructions

Contrary to the majority of studies investigating recall-to-reject, in the present work, participants were not informed about study and test list construction and were not given explicit recall-to-reject instructions (except Experiment 2). The evidence for recall-to-reject provided by Experiments 1 and 3 confirms the assumption that participants are capable of identifying the study and test structure themselves and spontaneously use recall-to-reject without being explicitly instructed to do so.

In Experiment 2 explicit recall-to-reject instructions were given in order to maximise the utilisation of the recall-to-reject process. However, data have to be interpreted with caution as the design of Experiment 2 additionally differed from that of Experiment 1 with respect to the non-cued control condition. But as the replacement of the control condition should only have an effect on cued items, differences between experiments with respect to non-cued items are likely to be due to the explicit recall-to-reject instructions. The finding that the false recognition effect was somewhat larger in Experiment 1 than in Experiment 2 indicates that instructions increased the base rate of recall-to-reject in the control condition. This is in line with previous research (Gallo, 2004; Lampinen et al., 2004; Rotello et al., 2000).

12.3 Mutual Exclusivity

The finding that category cueing eliminated the false recognition effect in Experiments 1 to 3 but not in Experiment 4 is in line with previous research about the importance of mutual exclusivity of stimuli. In all but the last experiment, associated distractors were similar to only one study item. Therefore, the recall of the studied item could be used to correctly reject the associated distractor. In Experiment 4, mutual exclusivity was excluded by including pairs of phonological associated items in the study list. To increase participants' awareness of the presentation of word pairs during study, pairs of phonologically associated items were additionally included in the test list. The finding, that category cueing did not reduce the false recognition effect in Experiment 4 indicates that the study list structure prevented participants from effectively using recall-to-reject.

However, it cannot be inferred from the present experiments that mutual exclusivity of stimuli is a necessary precondition for recall-to-reject. A reduction of the false recognition effect might sill have been observed if cued associated distractors were associated to two instead of only one item. If participants would be able to recall both associated items, then the associated distractor could be rejected if participants were aware that no more than two associated distractors were presented during study (Gallo, 2003, 2004). However, such a list structure would have required rhyme triples as stimulus material, with all items being exemplars of a different category (e.g., building, house; rodent, mouse; insect, louse). This would make list construction extremely difficult and therefore was not realised in Experiment 4. However, this prediction could be easily tested using semantic association as a similarity manipulation. Additionally, it would be interesting to manipulate the number of associated items presented during study. Corresponding to the probability of exhaustively recalling all associated items, the probability of recall-to-reject should decrease with increasing number of associated items.

Overall, the present experiments demonstrate that mutual exclusivity of stimuli can prevent the usefulness of recall-to-reject. However, it cannot be inferred from the present results that mutual exclusivity is a necessary precondition. Depending on the study list structure, recall-to-reject might still be used without mutual exclusivity.

Further studies are needed to test the importance of exhaustive recall and awareness of the number of associated items presented.

12.4 Methodological Implications

The results of the present work demonstrate that recall-to-reject is a memory editing mechanism that is not restricted to associative recognition and discrimination tasks but also occurs in simple item recognition tasks. Furthermore, the present results show that recall-to-reject can reduce or even eliminate the false recognition effect. However, as discussed with respect to the simultaneous familiarity increase that was observed in Experiments 1 and 2, it is suggested that the failure of many previously published studies to find evidence for recall-to-reject in recognition data can be explained by a simultaneous increase in familiarity. Thus, the results of the present studies emphasise the importance of including a control condition that is solely influenced by familiarity.

The interpretation of the results of Experiment 1 and 2 could be further confirmed by studying the effects of category cueing using multinomial modeling. For example, with the recently proposed simplified conjoint recognition model (Stahl & Klauer, in press) it could be confirmed that category cueing increases the probability of recall-to-reject as well as the familiarity of associated distractors when semantic association is used as similarity manipulation. An application of the model would only require the alteration of response options for the item question. Instead of the old/new recognition test, participants would have to answer with either *old*, *related* or *new*. Category cueing should increase the recall-to-reject parameter[19] as well as the familiarity parameter of associated distractors, and the increased familiarity for cued standard distractors should be reflected in a higher response bias parameter for cued compared to non-cued standard distractors. Furthermore, simplified conjoint recognition should indicate if the higher hit rate for cued compared to non-cued targets is based on an increase of the

[19] A thorough validation of the recall-to-reject parameter of the simplified conjoint recognition model would be of great interest. Up to date, the validation of this parameter is restricted to the target priming technique (Brainerd et al., 1995). Contrary to predictions, repetition of items during study did not increase the recall-to-reject parameter (Stahl & Klauer, in press).

probability of recall-to-accept or a familiarity increase. However, despite the great advantage that the parameters of multinomial models have in providing estimates of the probability of the assumed cognitive processes, some questions will remain open and other methods have to be used to provide the answers. For example, the simplified conjoint recognition model cannot be used to investigate whether explicit recall-to-reject instructions increase the base rate of recall-to-reject as was supposed in the discussion of Experiment 2. As participants are required to classify items as old, related or new, instructions always imply information about study and test list construction and explicit recall-to-reject instructions that are typically used make no sense with respect to the response options.

Category cueing was manipulated within subject in the present work. The simplified conjoint recognition model allows for investigating the effect of within-subject manipulations and parameter estimates of a single condition are based on behavioural data of the same subjects. However, parameter estimates of the conjoint recognition model are based on three different instruction conditions manipulated between subject. Bearing in mind that parameter estimates of the conjoint recognition model constitute a great amount of evidence for recall-to-reject (Brainerd et al., 2006; Brainerd et al., 1999; Brainerd et al., 2003; Odegard et al., 2005; Rotello, 2001), the success of the conjoint recognition model might be partly based on the design. Since interindividual differences increase the error variance in a between-subjects design, basically, the detection of an effect of a manipulation should be easier using a within than a between-subjects design. However, the opposite may be true for effects of recall-to-reject manipulations. In a within-subjects design, recall-to-reject manipulations could increase participants' awareness of this mechanism and therefore simultaneously increase the probability of base rate recall-to-reject in the control condition. This would reduce the difference between control and manipulation condition and consequently reduce the likelihood of finding an effect of the manipulation. Thus, it might be interesting for further studies to compare the effects of manipulations of within and between-subjects designs.

An alternative explanation for the reduction of false memories that has been repeatedly discussed is the possibility of a criterion shift (e.g., Heit et al., 2004; Rotello & Heit, 1999). It is argued that a reduced false alarm rate can be due to the adoption of a more conservative response criterion (Snodgrass & Corwin, 1988). However, due to the side effect of familiarity increase, the baseline level of false alarms was increased in Experiments 1 and 2 and the calculation of a response criterion parameter would suggest that response bias was more liberal in the cueing condition. However, within a single recognition test, the response criterion is set globally and not shifted on an item-by-item basis (Starns et al., 2006; Verde & Rotello, 2007). As category cueing was manipulated within subject and items were presented in a random order during test, a criterion shift is unlikely in the present experiments. Additionally, except for Experiment 2, non-cued items were preceded by an irrelevant category question, which should reduce the discriminability between cued and non-cued trials and therefore prevent a criterion shift. Furthermore, in the present experiments, false alarm rates were analysed in relative terms. The false recognition effect should thus not be influenced by response bias. This is equivalent to the method of subtracting the false alarm rate for standard distractors from the false alarm rate of associated distractors, that has been applied to correct for guessing (e.g., Brainerd et al., 1995; Gallo, 2003). However, the difference between the false alarm rates for standard and associated distractors might overestimate the effect of association because it assumes that false alarms to standard distractors are solely due to guessing. This procedure does not take into account the possibility of detecting low familiarity distractors as new (as discussed in Chapter 3.1.2 of the Introduction).

Finally, the present results have implications for the type of similarity manipulation and the type of manipulation used to influence the probability of recall-to-reject. Although category cueing resulted in a simultaneous familiarity increase and a tendency to answer consistently, these side effects were not inevitable. As Experiment 3 shows, possible side effects of category cueing could be successfully avoided by simply using phonological association as the similarity manipulation. It would be interesting to investigate the effects of category cueing for other forms of perceptual similarity and different stimulus materials such as pictures. Since memory representations for pictures are usually more distinctive than for words (Israel & Schacter, 1997), using category names as retrieval cues might be more effective and possible side effects might not occur even when semantic association is used. Finally, it would be an achievement to develop manipulations that facilitate recall-to-reject without running the risk of simultaneously increasing familiarity. For example, providing participants with accuracy feedback after each trial could increase participants' awareness of the recall-to-reject process and maximise participants' effort to use recall-to-reject. As accuracy feedback is not provided until participants have made their recognition decision, it should not result in any side effects that oppose recall-to-reject. When investigating the effect of accuracy feedback on recall-to-reject, it would be important to keep an eye on the response criterion, as accuracy feedback has been found to produce dynamic criterion shifts (Verde & Rotello, 2007).

12.5 Theoretical Implications

The results of the present work provide evidence for recall-to-reject and therefore are in line with dual-process theories of recognition memory. The results of the conditional analysis of Experiment 3 suggested that category cueing facilitates the rejection of associated distractors via increasing the recall of corresponding studied items. The false alarm rate for cued associated distractors was lower if the preceding category question had been accepted compared to if it had been rejected. However, an acceptance of a category question does not have to be based on the recall of the corresponding category exemplar but could also be due to high familiarity or guessing. Therefore, the assumption above could be further confirmed by requiring participants to recall the corresponding category exemplar. One would expect the false alarm rate for cued associated distractors to be extremely low after the corresponding studied item has been recalled. Furthermore, requiring participants to additionally rate their confidence regarding the item question could prove the assumption that the recall of a studied item will result in a high confidence rejection of the corresponding associated distractor.

It has to be taken into account whether the present results could alternatively be explained by single-process theories of recognition memory. A spreading activation model (Collins & Loftus, 1975) or the implicit associative response account (Underwood, 1965) can easily explain the semantic false recognition effect observed in the non-cued conditions of Experiments 1 and 2. However, these single-process models can explain the phonological false recognition effect observed in the non-cued conditions of Experiments 3 and 4 only with difficulty. One would have to assume that implicit associative responses are not restricted to semantic associates but are also made for phonological associates. Alternatively, it could be assumed that activation spreads through a semantic as well as a phonetic network. However, the elimination of the false recognition effect that was observed in the cued conditions of Experiments 1 and 2 clearly contradict the assumptions of both theories. The false alarm rate for cued associated distractors would have been expected to be higher than for cued standard distractors due to the semantic similarity between category exemplars.

Global memory models are able to alternatively explain some results that have been taken as evidence for the recall-to-reject process. For example, the test-pair similarity effect (compare Chapter 4.1) can alternatively be explained by MINERVA 2 (Hintzman, 1988). In a two-alternative forced-choice test, the hit rate was higher if a target was paired with a distractor similar to the target item compared to the hit rate of a distractor similar to another previously studied item. It could be argued that the simultaneous presentation of a target and the corresponding associated distractor draws participants' attention to distinguishing features, which facilitates recall-to-reject. However, MINERVA 2 can also account for the test-pair similarity effect. The prediction of the effect derives from the assumption that correlated inputs produce correlated outputs. Therefore, the retrieved activation of a target and distractor are correlated if the distractor is similar to the target (Hintzman, 2001). The correlation implicates a lower variance and as the probability to choose the distractor should decrease with decreasing variance, the correlation should result in a higher hit rate. Thus, the test-pair similarity effect can also be explained without the assumption of a recall-to-reject process.

Furthermore, the results of judgment of frequency studies that have also been taken as indirect evidence for recall-to-reject (compare Chapter 5.1) can be explained by the differentiation version of SAM (Shiffrin, Ratcliff, & Clark, 1990)[20]. This version of the

[20] Other competing explanations exist for the differentiation effect (e.g., see also McClelland & Chappell, 1998).

model allows similarity relations between stimuli to change with experience. It is assumed that extensive study of two highly similar stimuli increases discriminability and therefore reduces similarity between the stimuli (also see Hunt & McDaniel, 1993). This assumption can explain the finding that the judged frequency of an item can decrease or stay constant with increasing number of presentations of a similar item.

Moreover, global memory models can also explain some evidence that has been presented for the don't recall-to-reject process (Strack & Bless, 1994). The idea of Strack and Bless was to vary memorability of words by their salience. Most of the studied words were exemplars of the category *tools*, and only a few words were single exemplars from other categories. Participants showed lower false recognition of salient distractor items (non-tools) compared to the non-salient items (tools). Strack and Bless argued that salient items were correctly rejected more often based on the absence of recollection that contrasts the expected memorability. However, Rotello (1999) proposed an alternative familiarity-based explanation. She argued that global memory models could also predict the higher false alarm rate for non-salient items. As non-salient items are similar to several studied items and salient items are more dissimilar to studied items, global memory models predict salient items to be less familiar than non-salient items. The finding that recognition performance could neither be predicted by the judged salience nor the perceived memorability in a series of experiments was interpreted against the don't recall-to-reject explanation and in favour of global memory models.

Basically, if an effect can be explained by two different theories, the one which makes fewer assumptions should be preferred (Popper, 1959). One could argue that if an effect can be explained by dual-process theories as well as global memory models, the advantage of global memory models is that only one mechanism is assumed to underlie recognition decisions. However, the need of additional assumptions (as the differentiation assumption discussed with respect to judgments of frequency) challenges the parsimony advantage of global memory models (Diana, Reder, Arndt, & Park, 2006).

Global memory models cannot explain the results of the present work without further assumptions. When associated distractors are taken as retrieval cues, the resulting activation is predicted to be higher than the activation of standard distractors. This prediction is based on the assumption that retrieval cues will be compared to all memory traces. Thus, global memory models can easily explain the semantic false recognition effect observed in the non-cued conditions of Experiments 1 and 2 as well as the phonological false recognition effect observed in the non-cued conditions of Experiments 3 and 4. However, it is unclear how global memory models would formally describe the category question preceding cued items. If it is assumed that the category name that has not been studied leaves a trace in memory during test, the activation resulting when a cued associated distractor is used as a retrieval cue should still be higher than when a cued standard distractor is used as a retrieval cue, because the cued associated distractor is similar to the category name as well as its corresponding target. Hence, global memory models of recognition memory that are formalised as one-process models, cannot explain the results of the present work. However, as global memory models have also been developed for the domain of recall (e.g., Raaijmakers & Shiffrin, 1981) it might be possible to combine the assumptions and formalise a global memory model that includes a recall process in recognition memory. For example, the category question could be formalised as cued-recall within MINERVA 2 (Hintzman, 1986). It could be further assumed that the outcome of the cued-recall process is used as additional retrieval cue for the subsequent recognition decision. Then, if the outcome of the cued-recall and the test item form a compound-cue, the resulting activation for cued associated distractors should be lower than that for non-cued associated distractors based on distinguishing features included in the compound retrieval cue.

Overall, the present results show that recognition can be influenced by both familiarity and recall-like processes, as dual-process theories have posited. However, the results of the present work cannot be taken as confirmation of the assumptions of all dual-process theories. For example, although fuzzy trace theory has been very successful regarding the study of false memories and the recall-to-reject process, fuzzy trace theory has difficulties explaining the phonological false recognition effect observed for non-cued items in Experiment 3 and 4. False recognition for associated distractors is assumed to be based on a gist based similarity judgment. But as gist trace is thought to represent the memory for the meaning of an item, only a semantic false recognition effect can be explained. One might argue that the phonological similarity could also be extracted as the *gist* of an event. However, this should only occur if multiple phonologically associated items are presented during study, which was not the case in the present experiments. In Experiment 3 no associated items were presented during study and in Experiment 4 no more than two associated items were presented during study.

With respect to the conjoint recognition model, it was argued that "the most commonly studied forms of surface resemblance (e.g., rhyming) are not clean manipulations, and the resulting false-recognition effects may be due to semantic factors." (Brainerd et al., 1999 p. 167). It is suggested that prerecognition processes occurring before the identification of an item might be the semantic basis for phonological false recognition. During prerecognition, a cohort of word candidates is assumed to be activated as the sound of a spoken word unfolds (Carreiras, Ferrand, Grainger, & Perea, 2005; Wallace, Stewart, & Malone, 1995; Wallace, Stewart, Shaffer, & Wilson, 1998; Wallace, Stewart, Sherman, & Mellor, 1995). According to this, for example the presentation of *storm* should also increase the activation of *stork*. However, this explanation seems very unlikely for the phonological false recognition effect as the rhymes used in the present studies did not differ with respect to the last letters but with respect to the first letter (e.g., house, mouse). Additionally, the prerecognition process of cohort theory is assumed for spoken word recognition, but words were presented visually in the present studies. Thus, fuzzy trace theory does not provide a plausible explanation for the phonological false recognition effect.

12.6 Practical Implications

The avoidance of false memories is especially important in the forensic context, for example for eyewitness testimony, forensic interviews, police interrogations or eyewitness identifications. It has been suggested that the consideration of moderating factors of recall-to-reject in interview procedures, could reduce false memories via facilitating the recall-to-reject process (Brainerd & Reyna, 2002b). However, the application of manipulations that have been used to investigate recall-to-reject (as described in Chapter 5 of the Introduction) is limited to test phase manipulations and those manipulations that do not require knowledge of the event that happened (such as the target priming technique).

To maximise the use of recall-to-reject, interviews should occur as soon as possible after events and eyewitnesses should be given explicit instructions about the need for detailed information and they should be instructed only to report details that are remembered. The latter issue is part of the Cognitive Interview (Geiselman, Fisher, MacKinnon, & Holland, 1985; Köhnken, Milne, Memon, & Bull, 1999; Memon & Higham, 1999). Furthermore, the principle not to ask specific questions straight away, but to start with an open ended question and then continue to ask more and more specific questions based on the eyewitnesses answer corresponds to the idea of category cueing. The open ended question should cue recall of specific information and consequently prevent the occurrence of false memories with respect to the following specific details. However, the results of the present work indicate that this principle could result in a simultaneous familiarity increase, which opposes the effect of recall-to-reject. This conjecture is in line with the finding that the Cognitive Interview increases the number of correctly as well as falsely reported details (Köhnken et al., 1999). However, the increase of falsely reported details could also be due to other techniques of the interview procedure (Memon, Wark, Bull, & Koehnken, 1997).

Finally, recreating the context is a principle of the Cognitive Interview that should facilitate recall-to-reject. The effect of context on recall-to-reject could be investigated in future studies. Basically, as recall performance has been shown to be better when the test context matches the study context (Godden & Baddeley, 1975), the probability of recall-to-reject should thus depend on the type of context. However, the present work supposes that recreating the context simultaneously increases the familiarity (also see McKenzie & Tiberghien, 2004). This assumption is in line with the observation of contextual priming effects in item recognition that have been found for the hit rate of targets (Taylor & Juola, 1974) or with respect to response latencies (Macht & O'Brien, 1980).

Investigating the effect of context on recall-to-reject would be of great theoretical importance since precise predictions about the effect of context have been made by the global memory model termed ICE Theory (Murnane & Phelps, 1993, 1994; Murnane, Phelps, & Malmberg, 1999). Furthermore, the investigation of context effects with respect to recall-to-reject would be interesting against the background of inconsistent evidence of context effects in recognition (for an overview see Dougal & Rotello, 1999). For example, environmental context effects have been provided by Dalton (1993), but not by Godden and Baddeley (1980). Additionally, the distinction between environmental, internal and local context would have to be taken into account. For instance, the effect of local context on recall-to-reject could be investigated using a cueing manipulation. When pairs of words are studied as typical of an associative recognition task, the to-be-remembered word is either accompanied by the same or a different word as during encoding (Clark & Shiffrin, 1992; Winograd, Karchmer, & Russell, 1971). Unlike the cues used in the present work, the cues used in a context experiment would have to be studied along with the to-be-remembered item.

Overall, practical implications are limited by the fact that moderating factors of recall-to-reject during test are rare and that the effect of some moderating factors could be opposed by a simultaneous familiarity increase. Furthermore, research to date suggests that the effectiveness of recall-to-reject is limited to events that are mutually exclusive.

12.7 Conclusions

The present work provides evidence for the recall-to-reject process in item recognition and demonstrates its potential to reduce false recognition errors. Category cueing that has been shown to improve recall performance is successful in increasing recall-to-reject and consequently reducing the false recognition effect. Thus, retrieval cues found to increase conscious recollection in recall tasks can also be used in item recognition tasks. Therefore, the results show that recognition can be influenced by both familiarity and recall-like processes, as dual-process theories have posited.

Additionally, the present work investigated moderating factors of the recall-to-reject process. Most importantly, results showed that participants spontaneously use recall-to-reject without being explicitly instructed to do so. However, explicit instructions seem to increase the probability of recall-to-reject. Furthermore, the present work confirms the importance of study list structure for the use of recall-to-reject. Results demonstrate that the usefulness of recall-to-reject can depend on mutual exclusivity of stimuli.

Furthermore, the present work demonstrates that manipulations used to facilitate recall-to-reject can have side effects which oppose recall-to-reject. The simultaneous familiarity increase observed can explain why researchers often failed to find evidence for recall-to-reject in recognition data. This finding has important theoretical, methodological as well as practical implications. Future research is necessary to further disentangle the numerous cognitive processes underlying recognition memory. The combination of different methodological approaches–such as the analysis of recognition data and multinomial modelling–seems especially promising to further investigate the characteristics of the recall-to-reject process.

13 References

Anderson, J. R., Bothell, D., Lebiere, C., & Matessa, M. (1998). An integrated theory of list memory. *Journal of Memory and Language*, *38*, 341-380.

Anderson, J. R., & Bower, G. H. (1972). Recognition and retrieval processes in free recall. *Psychological Review*, *79*, 97-123.

Anisfeld, M. (1969). False recognition produced by semantic and phonetic relations under two presentation rates. *Psychonomic Science*, *17*, 366-367.

Anisfeld, M., & Knapp, M. (1968). Association, Synonymity, and Directionality in False Recognition. *Journal of Experimental Psychology*, *77*, 171-179.

Atkinson, R. C., & Juola, J. F. (1974). Search and decision processes in recognition memory. In R. D. Luce, R. C. Atkinson, D. H. Krantz, & P. Suppes (Eds.), *Contemporary developments in mathematical psychology: I*. Learning, memory and thinking. Oxford, England.

Bahrick, H. P. (1970). Two-phase model for prompted recall. *Psychological Review*, *77*, 215-222.

Batchelder, W. H., & Riefer, D. M. (1990). Multinomial processing models of source monitoring. *Psychological Review*, *97*, 548-564.

Batchelder, W. H., & Riefer, D. M. (1999). Theoretical and empirical review of multinomial process tree modeling. *Psychonomic Bulletin and Review*, *6*, 57-86.

Bayen, U. J., Murnane, K., & Erdfelder, E. (1996). Source discrimination, item detection, and multinomial models of source monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 197-215.

Brainerd, C. J., & Reyna, V. F. (2002a). Fuzzy-trace theory and false memory. *Current Directions in Psychological Science*, *11*, 164-169.

Brainerd, C. J., & Reyna, V. F. (2002b). Recollection rejection: How children edit their false memories. *Developmental Psychology*, *38*, 156-172.

Brainerd, C. J., Reyna, V. F., & Estrada, S. (2006). Recollection rejection of false narrative statements. *Memory*, *14*, 672-691.

Brainerd, C. J., Reyna, V. F., & Kneer, R. (1995). False-recognition reversal: When similarity is distinctive. *Journal of Memory and Language*, *34*, 157-185.

Brainerd, C. J., Reyna, V. F., & Mojardin, A. H. (1999). Conjoint recognition. *Psychological Review*, *106*, 160-179.

Brainerd, C. J., Reyna, V. F., Wright, R., & Mojardin, A. H. (2003). Recollection rejection: False-memory editing in children and adults. *Psychological Review*, *110*, 762-784.

Brainerd, C. J., Stein, L. M., & Reyna, V. F. (1998). On the development of conscious and unconscious memory. *Developmental Psychology*, *34*, 342-357.

Brainerd, C. J., Wright, R., Reyna, V. F., & Mojardin, A. H. (2001). Conjoint recognition and phantom recollection. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*, 307-327.

Brandt, M. (2001). *Modellbasierte Analyse globaler Gedächtnismodelle.* Lengerich: Pabst Science Publishers.

Carreiras, M., Ferrand, L., Grainger, J., & Perea, M. (2005). Sequential Effects of Phonological Priming in Visual Word Recognition. *Psychological Science. Vol 16(8) Aug 2005, 585 589.*

Clark, S. E., & Shiffrin, R. M. (1992). Cuing effects and associative information in recognition memory. *Memory and Cognition*, *20*, 580-598.

Coane, J. H., & McBride, D. M. (2006). The role of test structure in creating false memories. *Memory and Cognition*, *34*, 1026-1036.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, *82*, 407-428.

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

Dalton, P. (1993). The role of stimulus familiarity in context-dependent recognition. *Memory and Cognition*, *21*, 223-234.

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

Dewhurst, S. A. (2001). Category repetition and false recognition: Effects of instance frequency and category size. *Journal of Memory and Language*, *44*, 153-167.

Dewhurst, S. A., & Anderson, S. J. (1999). Effects of exact and category repetition in true and false recognition memory. *Memory and Cognition*, *27*, 664-673.

Dewhurst, S. A., & Farrand, P. (2004). Investigating the phenomenological characteristics of false recognition for categorised words. *European Journal of Cognitive Psychology*, *16*, 403-416.

Diana, R. A., Reder, L. M., Arndt, J., & Park, H. (2006). Models of recognition: A review of arguments in favor of a dual-process account. *Psychonomic Bulletin and Review*, *13*, 1-21.

Dodd, M. D., Sheard, E. D., & MacLeod, C. M. (2006). Re-exposure to studied items at test does not influence false recognition. *Memory*, *14*, 115-126.

Dodson, C. S., & Hege, A. C. G. (2005). Speeded retrieval abolishes the false-memory suppression effect: Evidence for the distinctiveness heuristic. *Psychonomic Bulletin and Review*, *12*, 726-731.

Dodson, C. S., & Schacter, D. L. (2001). "If I had said it I would have remembered it": Reducing false memories with a distinctiveness heuristic. *Psychonomic Bulletin and Review*, *8*, 155-161.

Dougal, S., & Rotello, C. M. (1999). Context effects in recognition memory. *American Journal of Psychology*, *112*, 277-295.

Ebbinghaus, H. (1964). *Memory: A contribution to experimental psychology*. Oxford, England: Dover.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175-191.

Gallo, D. A. (2003). Using recall to reduce false recognition. *Dissertation Abstracts International: Section B: The Sciences and Engineering*, *63*, 4947.

Gallo, D. A. (2004). Using Recall to Reduce False Recognition: Diagnostic and Disqualifying Monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*, 120-128.

Gallo, D. A., Bell, D. M., Beier, J. S., & Schacter, D. L. (2006). Two types of recollection-based monitoring in younger and older adults: Recall-to-reject and the distinctiveness heuristic. *Memory*, *14*, 730-741.

Gallo, D. A., Cotel, S. C., Moore, C. D., & Schacter, D. L. (2007). Aging Can Spare Recollection-Based Retrieval Monitoring: The Importance of Event Distinctiveness. *Psychology and Aging*, *22*, 209-213.

Gallo, D. A., Roberts, M. J., & Seamon, J. G. (1997). Remembering words not presented in lists: Can we avoid creating false memories? *Psychonomic Bulletin and Review*, *4*, 271-276.

Gallo, D. A., Weiss, J. A., & Schacter, D. L. (2004). Reducing false recognition with criterial recollection tests: Distinctiveness heuristic versus criterion shifts. *Journal of Memory and Language*, *51*, 473-493.

Geiselman, R. E., Fisher, R. P., MacKinnon, D. P., & Holland, H. L. (1985). Eyewitness memory enhancement in the police interview: Cognitive retrieval mnemonics versus hypnosis. *Journal of Applied Psychology*, *70*, 401-412.

Ghetti, S. (2003). Memory for nonoccurrences: The role of metacognition. *Journal of Memory and Language*, *48*, 722-739.

Ghetti, S. (2008). Rejection of false events in childhood: A metamemory account. *Current Directions in Psychological Science*, *17*, 16-20.

Ghetti, S., & Castelli, P. (2006). Developmental differences in false-event rejection: Effects of memorability-based warnings. *Memory*, *14*, 762-776.

Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review*, *91*, 1-67.

Godden, D. R., & Baddeley, A. D. (1975). Context-dependent memory in two natural environments: On land and underwater. *British Journal of Psychology*, *66*, 325-331.

Godden, D. R., & Baddeley, A. D. (1980). When does context influence recognition memory? *British Journal of Psychology*, *71*, 99-104.

Greene, R. L. (1989). On the relationship between categorical frequency estimation and cued recall. *Memory and Cognition*, *17*, 235-239.

Gruppuso, V., Lindsay, D. S., & Kelley, C. M. (1997). The process-dissociation procedure and similarity: Defining and estimating recollection and familiarity in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*, 259-278.

Hege, A. C. G., & Dodson, C. S. (2004). Why Distinctive Information Reduces False Memories: Evidence for Both Impoverished Relational-Encoding and Distinctiveness Heuristic Accounts. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*, 787-795.

Heit, E., Brockdorff, N., & Lamberts, K. (2004). Strategic processes in false recognition memory. *Psychonomic Bulletin and Review*, *11*, 380-386.

Hintzman, D. L. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review*, *93*, 411-428.

Hintzman, D. L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review*, *95*, 528-551.

Hintzman, D. L. (2001). Similarity, global matching, and judgments of frequency. *Memory and Cognition*, *29*, 547-556.

Hintzman, D. L., & Curran, T. (1994). Retrieval dynamics of recognition and frequency judgments: Evidence for separate processes of familiarity and recall. *Journal of Memory and Language*, *33*, 1-18.

Hintzman, D. L., Curran, T., & Oppy, B. (1992). Effects of similarity and repetition on memory: Registration without learning? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 667-680.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, *6*, 65-70.

Howe, M. L. (1998). When distinctiveness fails, false memories prevail. *Journal of Experimental Child Psychology*, *71*, 170-177.

Hunt, R. R., & McDaniel, M. A. (1993). The enigma of organization and distinctiveness. *Journal of Memory and Language*, *32*, 421-445.

Israel, L., & Schacter, D. L. (1997). Pictorial encoding reduces false recognition of semantic associates. *Psychonomic Bulletin and Review*, *4*, 577-581.

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

Jacoby, L. L., Toth, J. P., & Yonelinas, A. P. (1993). Separating conscious and unconscious influences of memory: Measuring recollection. *Journal of Experimental Psychology: General*, *122*, 139-154.

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin*, *114*, 3-28.

Jones, C. M., & Heit, E. (1993). An evaluation of the total similarity principle: Effects of similarity on frequency judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*, 799-812.

Jones, T. C. (2006). Editing (out) generated study words in a recognition exclusion task: Effects of response signal delay and generation procedure. *Memory*, *14*, 712-729.

Jones, T. C., & Atchley, P. (2002). Conjunction error rates on a continuous recognition memory test: Little evidence for recollection. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 374-379.

Jones, T. C., & Atchley, P. (2006). Conjunction Errors, Recollection-Based Rejections, and Forgetting in a Continuous Recognition Task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*, 70-78.

Jones, T. C., & Jacoby, L. L. (2001). Feature and conjunction errors in recognition memory: Evidence for dual-process theory. *Journal of Memory and Language*, *45*, 82-102.

Köhnken, G., Milne, R., Memon, A., & Bull, R. (1999). The cognitive interview: A meta-analysis. *Psychology, Crime and Law*, *5*, 3-27.

Lampinen, J. M., Meier, C. R., Arnal, J. D., & Leding, J. K. (2005). Compelling Untruths: Content Borrowing and Vivid False Memories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 954-963.

Lampinen, J. M., & Odegard, T. N. (2006). Editorial: Memory editing mechanisms. *Memory*, *14*, 649-654.

Lampinen, J. M., Odegard, T. N., & Neuschatz, J. S. (2004). Robust Recollection Rejection in the Memory Conjunction Paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition.*, *30*, 332-342.

Lampinen, J. M., Watkins, K. N., & Odegard, T. N. (2006). Phantom ROC: Recollection rejection in a hybrid conjoint recognition signal detection model. *Memory*, *14*, 655-671.

Lewandowsky, S. (1986). Priming in recognition memory for categorized lists. *Dissertation Abstracts International*, *47*, 1305.

Light, L. L., LaVoie, D., Valencia Laver, D., Albertson Owens, S. A., & Mead, G. (1992). Direct and indirect measures of memory for modality in young and older adults. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 1284-1297.

Lindsay, D. S., & Johnson, M. K. (2000). False memories and the source monitoring framework: Reply to Reyna and Lloyd (1997). *Learning and Individual Differences*, *12*, 145-161.

Lloyd, M. E. (2007). Metamemorial influences in recognition memory: Pictorial encoding reduces conjunction errors. *Memory and Cognition*, *35*, 1067-1073.

Macht, M. L., & O'Brien, E. J. (1980). Familiarity-based responding in item recognition: Evidence for the role of spreading activation. *Journal of Experimental Psychology: Human Learning and Memory*, *6*, 301-318.

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

Mannhaupt, H. R. (1983). German category norms for verbal items in 40 categories. Produktionsnormen für verbale Reaktionen zu 40 geläufigen Kategorien. *Sprache und Kognition*, *2*, 264-278.

Marsh, E. J., & Dolan, P. O. (2007). Test-induced priming of false memories. *Psychonomic Bulletin and Review*, *14*, 479-483.

Marsh, E. J., McDermott, K. B., & Roediger, H. L., III. (2004). Does test-induced priming play a role in the creation of false memories? *Memory*, *12*, 44-55.

McClelland, J. L., & Chappell, M. (1998). Familiarity breeds differentiation: A subjective-likelihood approach to the effects of experience in recognition memory. *Psychological Review*, *105*, 724-760.

McDermott, K. B., & Watson, J. M. (2001). The rise and fall of false recall: The impact of presentation duration. *Journal of Memory and Language*, *45*, 160-176.

McKenzie, W. A., & Tiberghien, G. (2004). Context effects in recognition memory: The role of familiarity and recollection. *Consciousness and Cognition: An International Journal*, *13*, 20-38.

Memon, A., & Higham, P. A. (1999). A review of the cognitive interview. *Psychology, Crime and Law*, *5*, 177-196.

Memon, A., Wark, L., Bull, R., & Koehnken, G. (1997). Isolating the effects of the cognitive interview techniques. *British Journal of Psychology*, *88*, 179-197.

Memon, A., & Young, M. (1997). Desperately seeking evidence: The recovered memory debate. *Legal and Criminological Psychology*, *2*, 131-154.

Murdock, B. B. (1982). A theory for the storage and retrieval of item and associative information. *Psychological Review*, *89*, 609-626.

Murnane, K., & Phelps, M. P. (1993). A global activation approach to the effect of changes in environmental context on recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*, 882-894.

Murnane, K., & Phelps, M. P. (1994). When does a different environmental context make a difference in recognition? A global activation model. *Memory and Cognition*, *22*, 584-590.

Murnane, K., Phelps, M. P., & Malmberg, K. (1999). Context-dependent recognition memory: The ICE theory. *Journal of Experimental Psychology: General*, *128*, 403-415.

Myers, J. L., & Well, A. D. (1995). *Research design and statistical analysis*. Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Neely, J. H. (1991). Semantic priming effects in visual word recognition: A selective review of current findings and theories. In D. Besner & G. W. Humphreys (Eds.), *Basic processes in reading: Visual word recognition* (pp. 264-336). Hillsdale, NJ, England: Lawrence Erlbaum Associates Inc 350 pp.

Odegard, T. N., & Lampinen, J. M. (2005). Recollection rejection: Gist cuing of verbatim memory. *Memory and Cognition*, *33*, 1422-1430.

Odegard, T. N., & Lampinen, J. M. (2006). Memory editing: Knowledge, criteria, and alignment. *Memory*, *14*, 777-787.

Odegard, T. N., Lampinen, J. M., & Toglia, M. P. (2005). Meaning's moderating effect on recollection rejection. *Journal of Memory and Language*, *53*, 416-429.

Payne, D. G., Elie, C. J., Blackwell, J. M., & Neuschatz, J. S. (1996). Memory illusions: Recalling, recognizing, and recollecting events that never occurred. *Journal of Memory and Language*, *35*, 261-285.

Popper, K. R. (1959). *The logic of scientific discovery*. Oxford, England: Basic Books.

Posner, M. I., & Keele, S. W. (1970). Retention of abstract ideas. *Journal of Experimental Psychology*, *83*, 304-308.

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

Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin*, *114*, 510-532.

Ratcliff, R., & McKoon, G. (1988). A retrieval theory of priming in memory. *Psychological Review*, *95*, 385-408.

Reyna, V. F., & Lloyd, F. (1997). Theories of false memory in children and adults. *Learning and Individual Differences*, *9*, 95-123.

Riefer, D. M., & Batchelder, W. H. (1988). Multinomial modeling and the measurement of cognitive processes. *Psychological Review*, *95*, 318-339.

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.

Rotello, C. M. (1999). Metacognition and memory for nonoccurrence. *Memory*, *7*, 43-63.

Rotello, C. M. (2001). Recall processes in recognition memory. In D. L. Medin (Ed.) *The psychology of learning and motivation: Advances in research and theory, Vol* (pp. 183-221). San Diego, CA, US: Academic Press.

Rotello, C. M., & Heit, E. (1999). Two-process models of recognition memory: Evidence for recall-to-reject? *Journal of Memory and Language*, *40*, 432-453.

Rotello, C. M., & Heit, E. (2000). Associative recognition: A case of recall-to-reject processing. *Memory and Cognition*, *28*, 907-922.

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.

Schacter, D. L., Israel, L., & Racine, C. (1999). Suppressing false recognition in younger and older adults: The distinctiveness heuristic. *Journal of Memory and Language*, *40*, 1-24.

Schacter, D. L., Verfaellie, M., Anes, M. D., & Racine, C. (1998). When true recognition suppresses false recognition: Evidence from amnesic patients. *Journal of Cognitive Neuroscience*, *10*, 668-679.

Scheithe, K., & Bäuml, K. H. (1995). German-language norms for representatives of 48 conceptual categories Deutschsprachige Normen fur Vertreter von 48 Kategorien. *Sprache und Kognition*, *14*, 39-43.

Seamon, J. G., Luo, C. R., Kopecky, J. J., Price, C. A., Rothschild, L., Fung, N. S., & Schwartz, M. A. (2002). Are false memories more difficult to forget than accurate memories? The effect of retention interval on recall and recognition. *Memory and Cognition*, *30*, 1054-1064.

Seamon, J. G., Luo, C. R., Schlegel, S. E., Greene, S. E., & Goldenberg, A. B. (2000). False memory for categorized pictures and words: The category associates procedure for studying memory errors in children and adults. *Journal of Memory and Language*, *42*, 120-146.

Shiffrin, R. M., Ratcliff, R., & Clark, S. E. (1990). List-strength effect: II. Theoretical mechanisms. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*, 179-195.

Smith, S. M., Tindell, D. R., Pierce, B. H., Gilliland, T. R., & Gerkens, D. R. (2001). The use of source memory to identity one's own episodic confusion errors. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*, 362-374.

Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General*, *117*, 34-50.

Sommers, M. S., & Lewis, B. P. (1999). Who really lives next door: Creating false memories with phonological neighbors. *Journal of Memory and Language*, *40*, 83-108.

Stahl, C., & Klauer, K. C. (in press). A Simplified Conjoint Recognition Paradigm for the Measurement of Gist and Verbatim Memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.

Starns, J. J., Hicks, J. L., & Marsh, R. L. (2006). Repetition effects in associative false recognition: Theme-based criterion shifts are the exception, not the rule. *Memory*, *14*, 742-761.

Strack, F., & Bless, H. (1994). Memory for nonoccurrences: Metacognitive and presuppositional strategies. *Journal of Memory and Language*, *33*, 203-217.

Taylor, G. A., & Juola, J. F. (1974). Priming effects on recognition performance. *Bulletin of the Psychonomic Society*, *3*, 277-279.

Tulving, E. (1981). Similarity relations in recognition. *Journal of Verbal Learning and Verbal Behavior*, *20*, 479-496.

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

Tulving, E., & Pearlstone, Z. (1966). Availability versus accessibility of information in memory for words. *Journal of Verbal Learning and Verbal Behavior*, *5(4)*, 381-391.

Tussing, A. A., & Greene, R. L. (1999). Differential effects of repetition on true and false recognition. *Journal of Memory and Language*, *40*, 520-533.

Underwood, B. J. (1965). False recognition produced by implicit verbal responses. *Journal of Experimental Psychology*, *70(1)*, 122-129.

Verde, M. F., & Rotello, C. M. (2007). Memory strength and the decision process in recognition memory. *Memory and Cognition*, *35*, 254-262.

Wallace, W. P., Malone, C. P., Swiergosz, M. J., & Amberg, M. D. (2000). On the generality of false recognition reversal. *Journal of Memory and Language*, *43*, 561-575.

Wallace, W. P., Stewart, M. T., & Malone, C. P. (1995). Recognition memory errors produced by implicit activation of word candidates during the processing of spoken words. *Journal of Memory and Language*, *34*, 417-439.

Wallace, W. P., Stewart, M. T., Shaffer, T. R., & Wilson, J. A. (1998). Are false recognitions influenced by prerecognition processing? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 299-315.

Wallace, W. P., Stewart, M. T., Sherman, H. L., & Mellor, M. D. (1995). False positives in recognition memory produced by cohort activation. *Cognition*, *55*, 85-113.

Whittlesea, B. W. A., & Leboe, J. P. (2003). Two fluency heuristics (and how to tell them apart). *Journal of Memory and Language*, *49*, 62-79.

Winer, B. J. (1971). *Statistical principles in experimental design* (2nd ed.). Tokyo: McGraw-Hill.

Winograd, E., Karchmer, M. A., & Russell, I. S. (1971). Role of encoding unitization in cued recognition memory. *Journal of Verbal Learning and Verbal Behavior*, *10*, 199-206.

Wortschatz Lexikon Deutsch. (2005). Universität Leipzig. Retrieved April 09, 2005, from http://wortschatz.informatik.uni-leipzig.de.

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

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

Appendix A

Item material used in Experiments 1 and 2 (108 category names with two exemplars each).

Category name	Exemplar 1	Exemplar 2
Alkoholisches Getränk [alcoholic beverage]	Wein [wine]	Sekt [sparkling wine]
Automarke [automobile brand]	Ford [Ford]	Fiat [Fiat]
Baumaschine [construction equipment]	Bagger [excavator]	Kran [crane]
Baumaterial [building material]	Zement [cement]	Mörtel [plaster]
Behälter [container]	Kasten [box]	Kiste [case]
Besteck [cutlery]	Gabel [fork]	Löffel [spoon]
Blasinstrument [wind instrument]	Trompete [trumpet]	Posaune [trombone]
Blume [flower]	Rose [rose]	Tulpe [tulip]
Bundesstaat der USA [state of the USA]	Utah [Utah]	Ohio [Ohio]
Christliches Fest [Christian Feast Day]	Pfingsten [Pentecost]	Ostern [Easter]
Edelmetall [precious metal]	Gold [gold]	Silber [silver]
Edelstein [precious stone]	Rubin [ruby]	Smargd [emerald]
Erkältungserscheinung [cold symptoms]	Husten [cough]	Schnupfen [cold]
Facharzt [medical specialist]	Kardiologe [cardiologist]	Neurologe [neurologist]
Fahrzeugteil [part of vehicle]	Gaspedal [accelerator]	Kupplung [clutch]
Familienfeier [family celebration]	Geburtstag [birthday]	Hochzeit [wedding]
Farbe [colour]	Grün [green]	Gelb [yellow]
Fisch [type of fish]	Forelle [trout]	Flunder [flounder]
Frauenname [female name]	Kerstin [Kerstin]	Katrin [Katrin]
Fußbodenbelag [floor covering]	Parkett [parquet]	Laminat [laminate]
Gartengerät [gardening tool]	Schaufel [shovel]	Spaten [spade]
Gebetshaus [place of worship]	Synagoge [synagogue]	Moschee [mosque]
Geflügel [poultry]	Huhn [chicken]	Henne [hen]
Gefühlszustand [emotion]	Glück [happiness]	Freude [enjoyment]
Geisteskrankheit [mental disease]	Neurose [neurosis]	Psychose [psychosis]
Geisteswissenschaft [humanities]	Philosophie [philosophy]	Theologie [theology]
Geistlicher [clergyman]	Pfarrer [pastor]	Priester [priest]
Gemüse [vegetable]	Tomate [tomato]	Paprika [pepper]
Geometrische Figur [geometrical shape]	Dreieck [triangle]	Quadrat [square]
Getreide [cereals]	Hafer [oats]	Gerste [barley]
Gewässer [bodies of water]	See [lake]	Teich [pond]
Gewichtsmaß [unit of weight]	Kilo [kilo]	Gramm [gram]
Gewürz [spice]	Pfeffer [pepper]	Salz [salt]
Gliedmaße [limbs]	Arm [arm]	Hand [hand]
Handwerker [craftsman]	Schreiner [joiner]	Schlosser [metalsmith]
Haustier [pet]	Hund [dog]	Katze [cat]
Heißgetränk [hot beverage]	Espresso [espresso]	Cappuccino [cappuccino]
Himmelsrichtung [cardinal direction]	Norden [north]	Süden [south]
Inneres Organ [internal organ]	Leber [liver]	Niere [kidney]
Insekt [insect]	Mücke [midge]	Moskito [mosquito]
Jahreszeit [season]	Sommer [summer]	Winter [winter]
Kampfsport [martial arts]	Judo [judo]	Karate [karate]
Kinderkrankheit [childhood illness]	Masern [measles]	Mumps [mumps]
Klassisches Musikstück [classical piece of music]	Symphonie [symphony]	Sonate [sonata]
Kleidungsstück [garment]	Hose [trousers]	Hemd [shirt]
Kontinent [continent]	Asien [Asia]	Afrika [Africa]
Kraftstoff [fuel]	Benzin [petrol]	Diesel [diesel]
Kräuter [herbs]	Petersilie [parsley]	Schnittlauch [chives]
Kücheneinrichtung [kitchen appliances]	Herd [stove]	Spüle [sink]
Kunstobjekt [work of art]	Statue [statue]	Skulptur [sculpture]
Land [country]	Schweden [Sweden]	Norwegen [Norway]
Landschaftsform [landscape]	Berg [mountain]	Tal [valley]
Laubbaum [deciduous tree]	Buche [beech]	Birke [birch]

Category name	Exemplar 1	Exemplar 2
Lebensabschnitt [period of life]	Jugend [youth]	Kindheit [childhood]
Lichtquelle [light source]	Laterne [lantern]	Lampe [lamp]
Längenmaß [unit of length]	Zentimeter [centimetre]	Millimeter [millimetre]
Männername [male name]	Markus [Markus]	Michael [Michael]
Märchengestalt [fairy tale character]	Zwerg [dwarf]	Riese [giant]
Menschenaffe [ape]	Gorilla [gorilla]	Schimpanse [chimpanzee]
Milchprodukt [dairy product]	Joghurt [yogurt]	Quark [curd cheese]
Militärischer Titel [military title]	General [general]	Offizier [officer]
Möbelstück [piece of furniture]	Sessel [armchair]	Sofa [sofa]
Monat [month]	Juni [June]	Juli [July]
Musikstil [style of music]	Jazz [Jazz]	Blues [Blues]
Musischer Beruf [musical occupation]	Dirigent [conductor]	Komponist [composer]
Nadelbaum [conifer]	Tanne [fir]	Fichte [spruce]
Nähzeug [sewing kit]	Nadel [needle]	Faden [thread]
Naturkatastrophe [natural disaster]	Erdbeben [earthquake]	Überschwemmung [flooding]
Naturwissenschaft [natural science]	Physik [physics]	Chemie [chemistry]
Nudelsorte [type of pasta]	Makkaroni [macaroni]	Tortellini [tortellini]
Obst [fruit]	Apfel [apple]	Birne [pear]
Optisches Gerät [optical instrument]	Teleskop [telescope]	Mikroskop [microscope]
Planet [planet]	Jupiter [Jupiter]	Venus [Venus]
Politisches Amt [government position]	Minister [minister]	Abgeordneter [assemblyman]
Postsendung [postal item]	Brief [letter]	Paket [parcel]
Printmedien [print media]	Zeitschrift [magazine]	Zeitung [newspaper]
Raubtier [predator]	Löwe [lion]	Tiger [tiger]
Rauschgift [narcotic drug]	Heroin [heroin]	Kokain [cocaine]
Rechenoperation [arithmetic operation]	Addition [addition]	Subtraktion [subtraction]
Reinigungsgerät [cleaning tool]	Putzlappen [cleaning cloth]	Staubsauger [vacuum cleaner]
Saiteninstrument [string instrument]	Gitarre [guitar]	Geige [violin]
Satzzeichen [punctuation mark]	Punkt [full stop]	Komma [comma]
Schreibgerät [writing utensil]	Bleistift [pencil]	Füller [pen]
Singvogel [singing bird]	Amsel [blackbird]	Drossel [thrush]
Sinnesorgan [sensory organ]	Nase [nose]	Ohr [ear]
Staatsform [form of government]	Demokratie [democracy]	Diktatur [dictatorship]
Stadt [city]	Hamburg [Hamburg]	München [Munich]
Süßigkeit [sweets]	Schokolade [chocolate]	Bonbon [candy]
Tabakware [tobacco products]	Zigarette [cigarette]	Zigarre [cigar]
Tageszeit [time of day]	Nacht [night]	Abend [evening]
Tanz [dance]	Samba [Samba]	Rumba [Rumba]
Teil des Auges [part of the eye]	Netzhaut [retina]	Pupille [pupil]
Textilart [types of textile]	Samt [velvet]	Seide [silk]
Tischlerwerkzeug [tool of a joiner]	Hobel [plane]	Säge [saw]
Verbrechen [crime]	Mord [murder]	Raub [robbery]
Verkehrsmittel [means of transportation]	Auto [car]	Bus [bus]
Verwandter [relative]	Neffe [nephew]	Nichte [niece]
Waffe [weapon]	Pistole [gun]	Gewehr [rifle]
Währung [currency]	Euro [Euro]	Dollar [Dollar]
Waschgelegenheit [washing facility]	Dusche [shower]	Badewanne [bathtub]
Wassersportart [aquatic sport]	Segeln [sailing]	Surfen [surfing]
Wetterphänomen [weather phenomenon]	Donner [thunder]	Blitz [lightning]
Wettkampfstätte [sports venue]	Arena [arena]	Stadion [stadium]
Wintersportart [winter sport]	Skifahren [skiing]	Eislaufen [ice scating]
Wochentag [day of the week]	Montag [Monday]	Dienstag [Tuesday]
Wohnmöglichkeit [type of residence]	Haus [house]	Hütte [hut]
Zeitmaß [unit of time]	Stunde [hour]	Minute [minute]
Zirkuskünstler [circus performer]	Akrobat [acrobat]	Artist [artist]

Appendix B

Item material used in Experiments 3 and 4. The item material consisted of a single category item pool (140 category names with one exemplar each) and a category pairs item pool (40 category pairs, i.e. two category names whose exemplars rhyme and were orthographically similar).

Single category item pool

Category name	Exemplar 1
Australisches Beuteltier [australian marsupial]	Känguruh [kangaroo]
Badeanstalt [public baths]	Freibad [open air bath]
Badebekleidung [bathing suit]	Bikini [bikini]
Ballsportart [ball game]	Tennis [tennis]
Bastelutensilie [craft material]	Kleber [glue]
Baumaschine [construction equipment]	Bagger [excavator]
Baustoff [building material]	Zement [cement]
Begrüßgungsgeste [gesture of greeting]	Umarmung [hug]
Beruf im Flugzeug [job in an aircraft]	Pilot [pilot]
Bildungsstätte [educational institution]	Schule [school]
Blasinstrument [wind instrument]	Flöte [flute]
Bundesstaat der USA [state of the USA]	Texas [Texas]
Christliches Fest [Christian Feast Day]	Ostern [Easter]
Delikatesse [delicacy]	Kaviar [caviar]
Deutsches Bundesland [federal state of Germany]	Bayern [Bavaria]
digitales Speichermedium [digital storage medium]	Diskette [floppy disk]
Edelgas [noble gas]	Helium [helium]
Edelmetall [precious metal]	Gold [gold]
Edelstein [precious stone]	Smaragd [emerald]
Entdecker [discoverer]	Kolumbus [Columbus]
Erkältungserscheinung [cold symptoms]	Husten [cough]
Europäisches Land [European country]	Schweiz [Switzerland]
Facharzt [medical specialist]	Chirurg [surgeon]
Familienfeier [family celebration]	Geburtstag [birthday]
Fast Food [fast food]	Pommes [potato chips]
Fernöstliches Land [far eastern country]	Japan [Japan]
Flächenmaß [square measure]	Hektar [hectare]
Flaschenverschluss [bottle cap]	Korken [cork]
Fußballbegriff [soccer terminology]	Abseits [offside]
Fußbodenbelag [floor covering]	Parkett [parquet]
Gastronomischer Beruf [gastronomical job]	Koch [cook]
Gaststätte [restaurant]	Restaurant [restaurant]
Gebirge [mountain range]	Alpen [Alps]
Geflügel [poultry]	Huhn [hen]
Geisteskrankheit [mental disease]	Psychose [psychosis]
Gemüse [vegetables]	Spinat [spinach]
Geometrische Figur [geometrical shape]	Kreis [circle]
Geräusch [sound]	Knall [bang]
Gesangsstimme [singing voice]	Tenor [tenor]
Gewichtseinheit [unit of weight]	Gramm [gram]
Gezeiten [tides]	Ebbe [low tide]
Glaubensbuch [religious text]	Koran [Koran]
Glücksspiel [game of chance]	Lotto [lottery]
Griechischer Buchstabe [greek letter]	Alpha [alpha]
Griechischer Gott [ancient greek god]	Zeus [Zeus]
Grundstücksbegrenzung [boundary around property]	Zaun [fence]

Category name	Exemplar 1
Haushaltstätigkeit [housekeeping task]	Bügeln [ironing]
Heimische Frucht [local fruit]	Apfel [apple]
Herrscher [monarch]	König [king]
Himmelsrichtung [cardinal direction]	Norden [north]
Hinrichtungsgerät [execution apparatus]	Galgen [gallows]
Hochschulabschluss [academic degree]	Magister [master's degree]
Insekt [insect]	Biene [bee]
Jahreszeit [season]	Herbst [autumn]
Kampfsport [martial arts]	Judo [judo]
Kartenspiel [card game]	Skat [Skat]
Kinderkrankheit [childhood illness]	Masern [measles]
Klassisches Musikstück [classical piece of music]	Sonate [sonata]
Kleiderverschluss [garment fastener]	Knopf [button]
Kletterpflanze [creeping plant]	Efeu [ivy]
Klimazone [climatic zone]	Tropen [tropics]
Knochen [bone]	Rippe [rib]
Kontinent [continent]	Afrika [Africa]
Körperflüssigkeit [bodily fluid]	Speichel [saliva]
Krankheitserreger [infectious agent]	Virus [virus]
Kriminelle Organisation [criminal organisation]	Mafia [mafia]
Kulturelle Veranstaltung [cultural event]	Konzert [concert]
Kulturepoche [cultural era]	Gotik [gothic era]
Kunsthandwerk [handicraft]	Töpfern [pottery]
Kunstobjekt [work of art]	Statue [statue]
Kuscheltier [cuddly toy]	Teddy [teddy bear]
Längenmaß [unit of length]	Meile [mile]
Laubbaum [deciduous tree]	Birke [birch]
Lebensabschnitt [period of life]	Jugend [youth]
Lebensgemeinschaft [alliance for life]	Ehe [marriage]
Lebensmittelgeschäft [grocery store]	Supermarkt [supermarket]
Luftfahrzeug [aircraft]	Hubschrauber [helicopter]
Medikament [medicine]	Aspirin [aspirin]
Medizinisches Instrument [medical instrument]	Skalpell [scalpel]
Menschenaffe [ape]	Gorilla [gorilla]
Meßgerät [measuring instrument]	Waage [scale]
Militärischer Rang [military title]	General [general]
Mittelmeerinsel [Mediterranean island]	Kreta [Crete]
Möbelhaus [furniture shop]	Ikea [Ikea]
Nahverkehrsmittel [local public transport]	Bus [bus]
Naturkatastrophe [natural disaster]	Erdbeben [earthquake]
Naturwissenschaft [natural science]	Chemie [chemistry]
Nobelpreisträger [Nobel laureate]	Einstein [Einstein]
Norddeutsche Stadt [city in northern Germany]	Bremen [Bremen]
Nordseeinsel [island in northern Germany]	Sylt [Sylt]
Nudelsorte [type of pasta]	Spaghetti [spaghetti]
Philosoph [philosopher]	Platon [Platon]
Planet [planet]	Venus [Venus]
Politisches Amt [government position]	Kanzler [chancellor]
Postsendung [postal item]	Brief [letter]
Printmedien [print media]	Zeitung [newspaper]
Ratespiel [guessing game]	Quiz [quiz]
Raumfahrt-Beruf [occupation in a space flight]	Astronaut [astronaut]
Rauschgift [narcotic drug]	Heroin [heroin]
Rechenoperation [arithmetic operation]	Addition [addition]
Reinigungsgerät [cleaning tool]	Besen [broom]
Saiteninstrument [stringed instrument]	Geige [violin]
Sanitäranlage [sanitary installation]	Klo [toilet]
Satzzeichen [punctuation mark]	Komma [comma]
Schalentier [shellfish]	Muschel [mussel]
Schlafbekleidung [sleeping garment]	Pyjama [pyjamas]
Schlaginstrument [percussion instrument]	Pauke [bass drum]

Category name	Exemplar 1
Schlangenart [type of snake]	Kobra [cobra]
Schmuckstück [piece of jewellery]	Brosche [brooch]
Schreinerwerkzeug [tool of a joiner]	Säge [saw]
Singvogel [singing bird]	Meise [tomtit]
Sprache [language]	Englisch [English]
Staatsform [form of government]	Diktatur [dictatorship]
Stachelpflanze [succulent plant]	Kaktus [cactus]
Südamerikanisches Land [country in South America]	Peru [Peru]
Süddeutsche Stadt [city in southern Germany]	München [Munich]
Süssungsmittel [sweetener]	Zucker [sugar]
Tabakware [tobacco products]	Zigarre [cigar]
Tasteninstrument [keyboard instrument]	Orgel [organ]
Teil der Kirche [part of a church]	Altar [altar]
Teil des Baums [part of a tree]	Laub [leaves]
Teil des Fahrrads [part of a bicycle]	Lenker [handlebar]
Teil eines Atoms [part of an atom]	Elektron [electron]
Teil eines Schiffs [part of a ship]	Anker [anchor]
Teil eines Wortes [part of a word]	Silbe [syllable]
Teilgebiet der Mathematik [field of mathematics]	Algebra [algebra]
Unterwäsche [underwear]	Slip [panties]
Verbrecher [criminal]	Dieb [thief]
Verunreinigung [dirt]	Fleck [stain]
Vorspeise [starter]	Suppe [soup]
Währung [currency]	Euro [Euro]
Waschmittel [detergent]	Seife [soap]
Wasserpflanze [aquatic plant]	Alge [algae]
Wassersportart [aquatic sports]	Segeln [sailing]
Weltreligion [major religion]	Islam [Islam]
Wettkampfstätte [sports venue]	Arena [arena]
Wintersportart [winter sport]	Eishockey [ice hockey]
Wochentag [day of the week]	Mittwoch [Wednesday]
Zirkuskünstler [circus performer]	Artist [artist]
Zugvogel [migratory bird]	Storch [stork]

Category pairs item pool

Category name 1	Exemplar 1	Category name 2	Exemplar 2
Ansiedelung [settlement]	Dorf [village]	Moorprodukt [moorland product]	Torf [peat]
Backware [bakery product]	Brot [bread]	Farbe [colour]	Rot [red]
Behälter [container]	Dose [box]	Blume [flower]	Rose [rose]
Besteck [cutlery]	Gabel [fork]	Erzählform [type of narrative]	Fabel [fable]
Comicfigur [cartoon character]	Schlumpf [smurf]	Fußbekleidung [footwear]	Strumpf [sock]
Epidemie [epidemic]	Pest [Black death]	Tierbehausung [animal dwelling]	Nest [nest]
Fleischgericht [meat dish]	Braten [roast]	Gartengerät [gardening tool]	Spaten [spade]
Fuhrwerk [cart]	Kutsche [carriage]	Spielplatz-Gerät [playground equipment]	Rutsche [slide]
Gefühlszustand [emotion]	Wut [anger]	Kopfbedeckung [headcovering]	Hut [hat]
Gehhilfe [twalking aid]	Krücke [crutch]	Flussübergang [river crossing]	Brücke [bridge]
Gesellschaftsschicht [social class]	Adel [nobility]	Nähzeug [sewing kit]	Nadel [needle]
Getreideart [type of cereal]	Roggen [rye]	Laufsportart [running sport]	Joggen [jogging]
Gewässer [body of water]	Bach [creek]	Gebäudeteil [part of a building]	Dach [roof]
Gewürz [spice]	Salz [salt]	Bierzutat [ingredient of beer]	Malz [malt]
Gliedmaße [limb]	Bein [leg]	Alkoholisches Getränk [alcoholic beverage]	Wein [wine]
Haustier [pet]	Hund [dog]	Teil des Gesichts [part of the face]	Mund [mouth]
Huftier [hoofed animal]	Pferd [horse]	Küchengerät [kitchen appliances]	Herd [stove]
Inneres Organ [internal organ]	Milz [spleen]	Textilart [types of textile]	Filz [felt]
Kirchliches Gebäude [religious building]	Dom [cathedral]	Europäische Hauptstadt [European capital]	Rom [Rome]
Kraftstoff [fuel]	Diesel [diesel]	Gesteinsart [type of rock]	Kiesel [pebble]
Märchengestalt [fairy tale character]	Fee [fairy]	Heißgetränk [hot beverage]	Tee [tea]
Meeressäugetier [aquatic mammal]	Wal [whale]	Geländeform [landscape feature]	Tal [valley]
Milchprodukt [dairy product]	Butter [butter]	Schiff [boat]	Kutter [cutter]
Monat [month]	Mai [May]	Fisch [fish]	Hai [shark]
Nadelbaum [conifer]	Fichte [spruce]	Verwandter [relative]	Nichte [niece]
Natürliche Lichtquelle [natural light source]	Sonne [sun]	Geistlicher Beruf [religious occupation]	Nonne [nun]
Niederschlagsform [type of precipitation]	Regen [rain]	Kirchliche Handlung [religious act]	Segen [blessing]
Raubkatze [big cat]	Löwe [lion]	Meeresvogel [seabird]	Möwe [sea gull]
Ruhestätte [resting place]	Grab [grave]	Gangart des Pferdes [horse gait]	Trab [trot]
Schreibgerät [writing utensil]	Füller [pen]	Handwerksberuf [tradesman]	Müller [miller]
Schriftstellerischer Beruf [type of author]	Dichter [poet]	Juristischer Beruf [juridical occupation]	Richter [judge]
Sitzmöbel [seating furniture]	Sofa [sofa]	Zweirad [two wheeler]	Mofa [moped]
Straftat [crime]	Mord [murder]	Automarke [automobile brand]	Ford [Ford]
Tageszeit [time of day]	Nacht [night]	Zahl [number]	Acht [eight]
Tanz [type of dance]	Tango [tango]	Tropische Frucht [tropical fruit]	Mango [mango]
Vergrößerungsgerät [magnification device]	Lupe [loupe]	Warngerät [alarm]	Hupe [horn]
Waffe [weapon]	Dolch [dagger]	Amphibie [amphibian]	Molch [newt]
Winterkleidung [winter garment]	Mantel [coat]	Trainingsgerät [training equipment]	Hantel [dumbbell]
Wohnmöglichkeit [type of residence]	Haus [house]	Nagetier [rodent]	Maus [mouse]
Zeiteinheit [unit of time]	Stunde [hour]	Verletzung [injury]	Wunde [wound]

Hiermit erkläre ich, dass ich die hier vorgelegte Dissertation eigenständig und ohne unerlaubte Hilfe angefertigt habe. Ich habe die Dissertation in der vorgelegten oder in ähnlicher Form noch bei keiner anderen Institution eingereicht. Ebenso versichere ich, dass ich bisher keine erfolglosen Promotionsversuche unternommen habe.

Düsseldorf, den 30.05.2008

(Juliane Schmid)