MEMORY, 2010, 18 (8), 863–882

Psychology Press
Taylor & Francis Group

Recall-to-reject: The effect of category cues on false recognition

Juliane Schmid[1], Sibylle C. Herholz[2], Martin Brandt[3], and Axel Buchner[1]

[1]Heinrich Heine University Düsseldorf, Germany
[2]Westfälische-Wilhelms-University Münster, Germany
[3]University of Mannheim, Germany

Four experiments examined the effect of category cueing on recall-to-reject, one of the central memory-editing mechanisms thought to prevent the occurrence of false memories. When category names were used as retrieval cues, the typically observed false recognition effect was eliminated for semantically associated distractors (Experiment 1a) and, moreover, a reduction in the absolute level of the false alarm rate was found for phonologically associated distractors (Experiment 2a). In addition to the old/new-recognition data, analyses using multinomial models support the interpretation that category cueing was successful in increasing the probability of recall-to-reject (Experiments 1b and 2b). The results are in line with dual-process theories of recognition memory and provide further evidence for recall-to-reject in single item recognition. They demonstrate its potential to reduce false recognition even when explicit instructions are not given. In addition, the results demonstrate that the paradigm can give rise to side effects that oppose recall-to-reject. A simultaneous familiarity increase can explain why many studies failed to find evidence for recall-to-reject in terms of false alarm rates.

Keywords: Memory editing; Recall-to-reject; False recognition; Category cues; Dual-process theories of recognition memory.

During the past decade there has been an increasing interest in memory editing mechanisms that can prevent the occurrence of false memories (Lampinen & Odegard, 2006). Recall-to-reject is one of the central memory editing mechanisms in recognition memory. A false recognition effect—that is, a systematically heightened false alarm rate for distractor items that are associated with studied items relative to that for distractor items that are not associated with any studied item—can usually be observed for semantic associations (e.g., Anisfeld & Knapp, 1968; Roediger & McDermott, 1995; Underwood, 1965), phonologic associations (e.g., Anisfeld, 1969; Sommers & Lewis, 1999), or orthographic associations of words (e.g., Hintzman & Curran, 1994). 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 the studied item *rat* is recalled, participants should reject the word *mouse* despite its high familiarity. This memory editing mechanism has been referred to in the literature as *recall-to-reject* (e.g., Rotello, 2001; Rotello & Heit, 2000) or *disqualifying recall-to-reject* (e.g., Gallo, 2004; Gallo, Bell, Beier, & Schacter, 2006). Some authors prefer the term *recollection rejection* (e.g., Brainerd,

Address correspondence to: Juliane Schmid, Institut für Experimentelle Psychologie, Heinrich-Heine-Universität, Universitätsstr.1, D-40225 Düsseldorf, Germany. E-mail: juliane.schmid@uni-duesseldorf.de

This research was partly supported by a doctoral scholarship of the Heinrich Heine University granted to Juliane Schmid. We thank D. A. Gallo, C. Brainerd, and T. C. Jones for their helpful comments on an earlier version of this article.

http://www.psypress.com/memory DOI:10.1080/09658211.2010.517756

Reyna, Wright, & Mojardin, 2003; Lampinen, Odegard, & Neuschatz, 2004) taking a more differentiated conception of recall into account. According to Brainerd and colleagues (e.g., Brainerd, Reyna, & Howe, 2009; Brainerd, Wright, Reyna, & Payne, 2002), both a recollective and a reconstructive component are involved in recall. The term recollection rejection refers to rejections based on the recollection process that yields vivid phenomenology. However, although the reconstructive process is known to be error-prone (e.g., Brainerd et al., 2003; Lampinen et al., 2004; Odegard & Lampinen, 2005), it may also lead to correct rejections of associated distractors. In the work presented here we do not distinguish between these different processes of recall. Thus in the following we use the term recall-to-reject referring to the phenomenon of recall-like processes leading to a rejection of associated distractors (also compare Odegard & Lampinen, 2006).

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 recall-like processes (for an overview, see Diana, Reder, Arndt, & Park, 2006; Mandler, 2008; Yonelinas, 2002). A review of the variety of existing dual-process theories is beyond the scope of the present paper. We only outline fuzzy trace theory (Brainerd & Reyna, 2002a; Brainerd, Reyna, & Kneer, 1995), because the simplified conjoint recognition model we used in the present work is based on this theory (Stahl & Klauer, 2008). Fuzzy trace theory has become increasingly popular, especially with respect to the investigation of false memories and recall-to-reject, but is not uncontroversial (e.g., Higham & Vokey, 2004; Jones, Brown, & Atchley, 2007). Central to the fuzzy trace theory is the assumption of two independent memory traces for studied items. The gist trace represents memory for the meaning of an item, whereas the verbatim trace represents memory for surface properties of the item such as perceptual details of its presentation. False recognition of associated distractors is assumed to result from gist-based similarity judgements. Furthermore, it is assumed that the retrieval of the verbatim trace of the corresponding learned item will result in the correct rejection of an associated distractor.

Evidence for recall-to-reject differs with respect to the methods employed. Analysing false alarm rates, a reduction of the false alarm rate for associated distractors is indicative for an increase in recall-to-reject (Brainerd et al., 1995). However, it has to be taken into account that a ‘‘null effect’’ or even increased false alarm rates for associated distractors could be obtained despite a successful increase in recall-to-reject as manipulations can simultaneously increase familiarity (Arndt & Jones, 2008). To overcome difficulties in interpreting false alarm rates, various other methods have been proposed to study recall-to-reject. First, multinomial processing tree models have frequently been used to measure the contribution of different cognitive processes underlying the observed categorical data (Batchelder & Riefer, 1999; Erdfelder et al., 2009; Riefer & Batchelder, 1988). The conjoint recognition model (Brainerd, Reyna, & Mojardin, 1999) and the simplified conjoint recognition model (Stahl & Klauer, 2008) allow memory and guessing processes to be disentangled and the probability of recall-to-reject to be estimated. Both models are based on assumptions made in fuzzy trace theory. The simplified conjoint recognition model that is used in the present study will be described in detail later. Second, different versions of extended receiver operating characteristic (ROC) analysis can be applied for the analysis of recall-to-reject in recognition tasks. These approaches are based on the assumption that recall-to-reject results in high confidence rejections (e.g., Lampinen et al., 2004; Lampinen, Watkins, & Odegard, 2006; Rotello, 2001; Rotello, Macmillan, & Van Tassel, 2000). Finally, different types of self-reports have been used to estimate the occurrence of recall-to-reject (e.g., Gallo, 2004; Jones, 2005; Jones & Atchley, 2006; Lampinen et al., 2004; Odegard & Lampinen, 2005; Odegard, Lampinen, & Toglia, 2005).

Furthermore, studies on recall-to-reject differ with respect to the manipulations 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 recall of the corresponding learned item of the associated distractor under test. However, the outcome of recall manipulations depends on the type of recall component (recollective vs reconstructive) that is affected (e.g., Brainerd et al., 2002, 2009). Thus, no substantial increase of recall-to-reject is to be expected if a manipulation mainly increases the error-prone reconstructive component.

Many study phase manipulations have been used to influence the probability of recall-to-reject. These include manipulations of the length of study lists (e.g., Odegard et al., 2005), the level of attention (e.g., Jones & Jacoby, 2001; Lampinen et al., 2006; Odegard, Koen, & Gama, 2008; Odegard & Lampinen, 2005), the retention interval (e.g., Brainerd et al., 2003; Jones & Atchley, 2002, 2006; Odegard et al., 2005), or the number of item repetitions (e.g., Brainerd, Reyna, & Estrada, 2006; Brainerd et al., 1999; Hall & Kozloff, 1970; Jones, 2005; Jones & Jacoby, 2001; Lampinen et al., 2004; Leding & Lampinen, 2009; Stahl & Klauer, 2008; Tussing & Greene, 1999). Overall, evidence for recall-to-reject has been numerous in terms of model parameters and self-reports but is less frequent in terms of old/new recognition. For example, presenting targets three times during study compared to only once increased the recall-to-reject parameter of an 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). Compared to study phase manipulations, test phase manipulations known to affect recall are rare. The target priming technique—that is, presenting targets just prior to their associated distractors (Brainerd et al., 1995)—has been found to increase the recall-to-reject parameter of conjoint recognition (Brainerd et al., 1999; Brainerd, Stein, & Reyna, 1998) and simplified conjoint recognition (Stahl & Klauer, 2008). It even turned out to be a simple and powerful method to reduce false recognition errors (Brainerd et al., 1995; but not Brainerd et al., 1998), although its effectiveness seems to be partly due to response strategies (Wallace, Malone, Swiergosz, & Amberg, 2000).

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. Hence presenting category names seems to improve recall of the corresponding learned category exemplars (also see Epstein, Dupree, & Gronikowski, 1979; Hudson & Austin, 1970). Gallo (2004) had the idea to use this manipulation in a recognition task. In a pilot study he presented category names along with category exemplars during study and/or test. However, although the presentation of category names did improve recall, no effect on recognition was observed. In subsequent experiments he used category cueing and found evidence for recall-to-reject. The effect was strongest when participants were instructed to use recall-to-reject and the length of the category was short and known to the participants. These results imply that category cueing was important for the use of recall-to-reject. However, the effect of the cueing manipulation cannot be estimated because no effect of cueing on false recognition data was observed in the pilot study and the following experiments did not include a condition without cueing. We believe that category cueing is a clever manipulation that increases recall-to-reject and thus might also be used to reduce false recognition.

Therefore the present work aimed at investigating the effect of cueing on recall-to-reject, and for that purpose extends Gallo's work (2004) by applying multinomial modelling and by changing some of the design characteristics. The first change was that, in our task, category names were never presented during study, and 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 during test. The second change was that the number of exemplars learned per category was restricted to only one to maximise the effectiveness of 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). Another difference concerns the instructions given. The majority of studies with evidence of recall-to-reject used explicit instructions about how the recall of studied items can be used to avoid false memories for associated distractors (e.g., Arndt & Jones, 2008; Gallo, Cotel, Moore, & Schacter, 2007; Lampinen et al., 2004; Odegard & Lampinen, 2005; Rotello, 2001). In contrast, the picture is much less clear with respect to studies without such explicit instructions: some failed to provide evidence for recall-to-reject (e.g., Rotello & Heit, 1999; Tussing & Greene, 1999), whereas some did report such evidence (Gallo, 2004; Hall & Kozloff, 1970). Nevertheless, when both explicit and no explicit instructions were compared, explicit instructions increased the likelihood of recall-to-reject (e.g., Gallo, 2004; Lampinen et al., 2004; Rotello et al., 2000). Given this, the need for more research on the role of instructions has been emphasised (Lampinen et al., 2006; Odegard & Lampinen, 2006). We thus reasoned that knowing whether and under what conditions participants spontaneously use recall-to-reject would be interesting, in that it indicates the relevance of this memory editing mechanism for everyday memory. Therefore we aimed at investigating the effect of cueing on recall-to-reject without giving any instructions about the nature of study and test list construction (except for modelling experiments) and without explicitly instructing participants to use recall-to-reject.

EXPERIMENT 1A

The purpose of the first experiment was to investigate whether category names, when presented as retrieval cues, can be used to reduce the false recognition effect in an old/new recognition paradigm without explicit recall-to-reject instructions. A single item recognition task with a semantic similarity manipulation at test was developed. Participants first learned a list of category exemplars. During the test phase participants responded to sequences of category and item questions. Examples of category questions would be ‘‘Was a *beverage* presented during study?’’ and ‘‘Was a *fruit* presented during study?’’ in the control and cued conditions, respectively. ‘‘Was the word *apple* presented during study?’’ would be an example of an item question. Thus category and item were unrelated in the control condition but were associated in the cued condition.

In a control condition a typical false recognition effect was expected. The false alarm rate was anticipated to be significantly higher for semantically associated distractors than for standard distractors. We assumed that presenting questions with category names would facilitate recall-to-reject. Thus the false recognition effect was expected to be reduced in the cued condition compared to the control condition. However, it seems likely that the categorical overlap between category and item question in the cued condition (e.g., fruit–apple/pear) leads to a general increase in familiarity that should apply to all item types (i.e., cued targets, cued distractors, and cued associated distractors). 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 item familiarity and thus subsequent false recognition. 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–recognise strategy (Anderson & Bower, 1972; Bahrick, 1970). Thus false recognition of the subsequent item could be regarded as a kind of source-monitoring failure because false recognition could be avoided if the increased familiarity of category exemplars were to be correctly 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 (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 underlying process, such a familiarity increase should raise the probability of ‘‘old’’ responses for all item types. An increase in recall-to-reject therefore does not necessarily have to result in an absolute reduction of the false alarm rate for associated distractors. Even an increased false alarm rate or a ‘‘null effect’’ could be indicative of recall-to-reject (Arndt & Jones, 2008). The crucial comparison is that of the false recognition effect between cued and control condition.

The hit rate is predicted to be higher for cued targets than for control targets. However, hit rates do not allow any conclusions to be drawn about the underlying mechanisms, as familiarity and recall should both increase the hit rate.

Method

Participants. Participants were 74 adults (mostly students), 46 of whom were female. The mean age of participants was 26 (between 17 and 50) years. All participants were native German speakers. Participants were tested in groups of up to four.

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 terms of the number of syllables, dominance (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).

Participants studied category exemplars and were later tested for recognition memory. An example of all item types is shown in Table 1. Associated stimuli were mutually exclusive; that is, only one exemplar per category was presented during study. 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., *apple* when *pear* was studied). 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.

During test, category names were presented as retrieval cues prior to the items. To guarantee that participants were attending to the retrieval cues and to maximise their effectiveness, cues were presented in the form of category questions: Participants had to decide whether an exemplar of this category had been in the study list or not. For instance, if *pear* was learned and *apple* was presented as the associated distractor, participants were asked in the preceding category question whether a *fruit* was learned. Please note that in contrast to the target priming technique (Brainerd et al., 1995) associated distractors' corresponding learned items (*pear* in the example above) were not tested.

In order to have an appropriate control condition, all items of the control condition were also preceded by a category question, but it pertained to an irrelevant category (e.g., *beverage?* when the following item is *apple*). For one half of the items within the control condition, a category exemplar of the irrelevant category question (e.g., *beer*) was presented during study.

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 category exemplars. All study list items 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 that were all preceded by a category question (relevant category question for cued items and irrelevant category question for control items). All items of the test list were presented in random order. Each test list was composed of four cued/control targets, cued/control distractors, and cued/control associated distractors respectively.

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 that comprised one test item of each item condition (compare Table 1) 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 instructed to memorise words as well as possible. They were not informed about details of study or test list construction and no explicit recall-to-reject instructions were given. A 30-second interval filled with simple arithmetic problems separated study and test phase.

TABLE 1
Examples for the different item conditions in Experiments 1a and 1b

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

All control items were 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 control items.

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. In the subsequent item recognition task a category exemplar was presented and participants had to decide whether this exemplar had been in the study list or not. 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 a study phase, a retention interval, and a test phase. Blocks were separated by a 1-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.

Design. Experiment 1a employed a two-factorial design with item type (target, distractor, associated distractor) and cueing condition (cued vs control) as within-participant variables. This resulted in six different item conditions, which are termed cued/control targets, cued/control distractors, and cued/control associated distractors. The dependent variable was the probability of old responses to the different item types.

All power calculations were conducted using G*Power 3 (Faul, Erdfelder, Lang, & Buchner, 2007). The false recognition effect was defined as the difference in false alarm rates for associated distractors and for standard distractors. As our main hypotheses refer to binary comparisons of recognition performance between specific item conditions, an analysis of variance as well as follow-up *t*-tests were computed. Given a sample size of $N = 74$ and $\alpha = .05$, it was possible to detect a false recognition effect size $d_z = 0.5$ with a probability of $1\text{-}\beta = .99$. 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 analyses, respectively (see Results for details).

Results

Participants responded correctly to 71% of the category questions ($SE = 1.05$) and to 86% of the item questions ($SE = 0.94$). The proportions of old responses to the different item types are shown in Table 2.

Proportions of old responses were analysed using a 3 (item type) × 2 (cueing condition) repeated measures analysis of variance. The

TABLE 2
Proportion of hits and false alarms for the different item conditions in Experiment 1a and 2a

		Unconditional	*Conditional*			
Experiment	*Item condition*	*M (SE)*	*Preceding category accepted M (SE)*	*Preceding category rejected M (SE)*	*t (df)*	d_z
1a	Control targets	.80 (.02)	.81 (.02)	.80 (.03)	0.07 (71)	0.01
	Control distractors	.05 (.01)	.04 (.01)	.06 (.01)	1.68 (71)	0.20
	Control associated distractors	.14 (.01)	.15 (.02)	.14 (.02)	0.36 (72)	0.04
	Cued targets	.84 (.02)	.92 (.02)	.63 (.04)	7.26 (64) *	0.91
	Cued distractors	.14 (.02)	.27 (.04)	.07 (.02)	5.91 (68) *	0.72
	Cued associated distractors	.14 (.02)	.15 (.02)	.12 (.03)	0.89 (65)	0.11
2a	Control targets	.81 (.01)	.80 (.02)	.83 (.02)	1.81 (63)	0.23
	Control distractors	.11 (.01)	.12 (.02)	.10 (.02)	0.89 (63)	0.11
	Control associated distractors	.15 (.02)	.16 (.02)	.16 (.02)	0.26 (63)	0.03
	Cued associated distractors	.13 (.02)	.10 (.02)	.19 (.03)	4.04 (63) *	0.51

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. * $p < .001$.

main effect of the item type was significant, $F(2, 72) = 560.02$, $p < .001$, $\eta^2 = .94$. Helmert contrasts revealed that the hit rate for targets ($M =$ .82) was higher than the false alarm rate for associated distractors and standard distractors ($M = .12$), $F(1, 73) = 1110.73$, $p < .001$, $\eta^2 = .94$, and that the false alarm rate for associated distractors ($M = .14$) was higher compared to standard distractors ($M = .09$), $F(1, 73) = 16.58$, $p < .001$, $\eta^2 = .19$; that is, a typical false recognition effect was observed. The main effect of the cueing condition was also significant, $F(1, 73) =$ 12.86, $p = .001$, $\eta^2 = .15$, showing more old responses in the cued condition ($M = .37$) compared to the control condition ($M = .33$). The main effects were qualified by a significant interaction, $F(2, 72) = 7.63$, $p = .001$, $\eta^2 = .18$. Follow-up *t*-tests indicate that, in the control condition, the false alarm rate for associated distractors ($M = .14$) was significantly higher than that for standard distractors ($M = .05$); that is, a typical false recognition effect was observed, $t(73) = 6.15$, $p < .001$, $d_z =$ 0.71. In the cued condition, however, the false alarm rate for associated distractors ($M = .14$) and standard distractors ($M = .14$) did not differ, $t(73) = 0.41$, $p = .683$, $d_z = 0.05$; that is, category cueing eliminated the false recognition effect. A closer look at the absolute false alarm rates reveals that the false alarm rate for cued standard distractors (.14) was higher than for control standard distractors (.05), $t(73) = 4.68$, $p < .001$, $d_z = 0.54$, and the false alarm rate for associated distractors did not differ between cued and control distractors (.14 respectively), $t(73) = 0.45$, $p =$ .654, $d_z = 0.05$. The hit rate for cued targets was higher ($M = .84$) than for control targets ($M =$ .80), $t(73) = 2.43$, $p = .009$, $d_z = 0.28$.

Conditional analysis. To obtain further information on why the false recognition effect was eliminated for cued items, the proportions of old responses were analysed depending on the response to the preceding category question. The category question preceding control items was unrelated to the subsequent item. Acceptance or rejection of the preceding category question was thus not expected to have any effect on the subsequent item recognition decision. In contrast, the category question preceding cued items was semantically related to the subsequent item. Item recognition decisions for cued items were thus likely to depend on the response to the preceding category question. For cued associated distractors, the false alarm rate was expected to be lower given that the preceding category question is accepted. This was thought to result from recall-to-reject because if the preceding category question had been accepted based on recall of the corresponding studied category exemplar, the discrepancy to the following associated distractor was likely to be detected and the associated distractor should have been rejected. For cued standard distractors, the conditional false alarm rates were expected to not differ because acceptance of a category question cueing a distractor was assumed to be based on guessing. Recognition of target items should be very high after the corresponding category question had been accepted (due to either recall or familiarity). Therefore the hit rate was expected to be higher after acceptance of the corresponding category question.

The results of the conditional analysis are presented in the right-hand side of Table 2. As expected, the response (yes or no) to the unrelated category question preceding control items affected neither the hit rate for targets (.81 vs .80), $t(71) = 0.07$, $p = .946$, $d_z = 0.01$, nor the false alarm rates for both standard distractors (.04 vs .06), $t(71) = 1.68$, $p = .098$, $d_z = 0.20$, and associated distractors (.15 vs .14), $t(72) = 0.36$, $p =$ .717, $d_z = 0.04$. However, contrary to predictions, the false alarm rate for cued standard distractors was significantly higher when the preceding category question had been accepted (.27) than when it had been rejected (.07), $t(68) = 5.91$, $p <$ .001, $d_z = 0.72$. The false alarm rate for cued associated distractors was not affected by the response to the preceding category question (.15 vs .12), $t(65) = 0.89$, $p = .379$, $d_z = 0.11$. The hit rate for targets was higher after acceptance of the corresponding category question (.92 vs .63), $t(64) = 7.26$, $p < .001$, $d_z = 0.91$.

Discussion

Cueing items by the corresponding category name not only reduced but even eliminated the false recognition effect. This indicates that category cueing was successful in increasing the probability of recall-to-reject. Thus the present experiment provides evidence for the assumption that participants are capable of identifying the test structure themselves and of using recall-to-reject spontaneously.

However, the absolute false alarm rate for cued associated distractors was not reduced compared to that for control associated distractors. The overall pattern of data suggests that this was due to a general increase in familiarity caused by the corresponding category questions. A familiarity increase as a side effect of category cueing should affect all cued item types, leading to a generally increased probability of old responses. This was indeed true for targets and standard distractors, but not for associated distractors, the latter reflecting the contribution of recall-to-reject. Given that both processes—recall-to-reject and familiarity—work in opposition, the absolute false alarm rate for associated distractors is neither increased nor reduced. This result is consistent with the suggestion that a ''null effect'' or even increased false alarm rates could be obtained despite a successful increase in recall-to-reject as manipulations could increase both recall and familiarity (Arndt & Jones, 2008).

The conditional analysis—that is, the analysis of the responses to the item questions depending on the responses to the preceding category question—shows that, as expected, it did not make a difference for all control item types whether the preceding unrelated category question had been accepted or rejected. In contrast, the false alarm rate for cued standard distractors was higher given that the preceding category question had been accepted. This difference was not expected and can only be based on strategic guessing. This indicates that participants tended to respond in such a manner that their response to a particular item question was consistent with their response to the preceding category question. Although such a response strategy should affect all cued item types, no difference was found for cued associated distractors. In fact, an opposite pattern as that found for standard distractors was expected for associated distractors due to recall-to-reject. Recall-to-reject is most likely 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 did indeed employ a strategy of responding consistently, the opposite pattern would be expected. Hence both processes work in opposition and seem to have cancelled each other out. The increased hit rates of targets in the cued compared to the control condition as well as after acceptance of the corresponding category question are in line with expectations but do not allow any conclusions to be drawn about the contribution of the underlying processes discussed.

To summarise, the results of Experiment 1a 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 recall-to-reject. However, the semantic similarity between category names and the corresponding exemplars simultaneously increased the familiarity of cued exemplars in general. Due to this side effect the baseline of false alarms was higher compared to control items and therefore no absolute reduction of the false alarm rate could be observed. Furthermore, the obvious semantic relation between category question and the following category exemplar likely encouraged a tendency to respond consistently; that is, if participants accepted that an item from a specific category had occurred during the learning phase, then they also accepted the subsequently presented item from this category as old. This second side effect obscures the pattern expected in the conditional analysis. Given that recall-to-reject as well as the processes assumed to underlie the observed side effects affect recognition data, additional methods such as multinomial models are needed to decompose observed performance into the underlying processes (Erdfelder et al., 2009).

EXPERIMENT 1B

The main purpose of Experiment 1b was to use the simplified conjoint recognition model (Stahl & Klauer, 2008) to decompose observed performance into the processes assumed to take place in the category cueing paradigm developed in Experiment 1a. The simplified conjoint recognition model assumes that participants study items and later complete a recognition test containing the three item types: targets, standard distractors, and associated distractors. Participants identify an item as old or new as in typical old/new recognition tasks, or as related. They are instructed to respond ''related'' if they believe an item to be unstudied but related to a studied item. This is similar to a test procedure introduced by Jones (2005) using the memory conjunction paradigm, where participants were required after old/new judgements to indicate for items identified as new if rejection was based on recall of the parent item or low familiarity (remember vs know response, also see Arndt & Jones, 2008; Jones & Atchley, 2006).

With respect to behavioural data it is difficult to make precise predictions about the effect of category cueing on the different response categories. In terms of recall-to-reject the related response category seems most interesting. An increase of recall-to-reject in the cued condition compared to the control condition should be reflected in an increase in related responses for associated distractors. Without any side effects, the new responses for standard distractors should not differ between conditions. However, an increased familiarity that was observed in Experiment 1a should lower the new responses in the cued condition. But it is difficult to predict how this difference divides into old and related responses. Thus an increase in related responses for associated distractors does not necessarily have to be due to recall-to-reject but could be partially based on an increase in familiarity. For targets, an increase in old responses should be observed in the cued condition, based on an increase in familiarity and potentially recall.

The conditional analysis of behavioural data should not show any differences of response frequencies for all control items after acceptance or rejection of the preceding category question. Due to recall-to-reject, the proportion of related responses for cued associated distractors is expected to be higher after acceptance of the preceding category question. However, this difference could partly be caused by a strategy of responding consistently. This side effect should be reflected in fewer new responses for cued standard distractors after acceptance of the category question. A higher proportion of old responses for cued targets after acceptance of the category question could be based on an increase in recall or a strategy to respond consistently or both.

Obviously the contribution of different types of processes to observed performance makes it difficult to derive precise predictions with respect to behavioural data. Therefore we used the simplified conjoint recognition model (Stahl & Klauer, 2008) to test our predictions about the processes underlying memory performance. An adaptation of the model for the present purposes is presented in Figure 1.

The model in Figure 1 contains six parameters. Each parameter represents the probability with which certain cognitive processes occur. When a target is presented at test, parameter V_t in the upper model tree represents the probability of retrieving a target's verbatim trace, which results in a correct identification as "old". Given no retrieval of the verbatim trace (1-V_t), the parameter G_t represents the probability of retrieving a target's gist trace. In this case participants may still guess, with probability a, that the item is "old", or, with probability 1-a, that the item is "related". If neither the verbatim nor the gist trace can be retrieved participants may still guess, with probability b, that the item is either old or related (and again decide with probability a that the item is "old"), or, with probability 1-b, that the item is "new".

The middle tree diagram in Figure 1 represents the processes assumed to occur when associated distractors are presented at test. The recall-to-reject parameter corresponds to the parameter V_r, which represents the probability that the verbatim trace of the corresponding learned item is retrieved. This results in an identification of the associated distractor and, consequently, in a correct classification as "related". If the corresponding learned item's trace is not retrieved, which occurs with probability 1-V_r, then parameter G_r represents the probability of retrieving the corresponding learned item's gist trace. Identical to the upper tree diagram, participants then may still guess, with probability a, that the item is "old", or, with probability 1-a, that the item is "related". Given that neither the verbatim nor the gist trace can be retrieved, the sequence of the processes a and b for associated distractors is the same as for targets as well as for standard distractors (compare the upper and the lower tree diagrams in Figure 1). The lower tree diagram illustrates that responses to standard distractors are assumed to rely only on processes that involve neither gist nor verbatim memory traces. These guessing processes are represented by parameters a and b.

To test the effect of category cueing, two sets of the three model trees depicted in Figure 1 are needed, one for the cued and another for the control condition. The parameters of the model can be interpreted in terms of probabilities with which the cognitive processes that they represent occur. If category cueing increases the probability

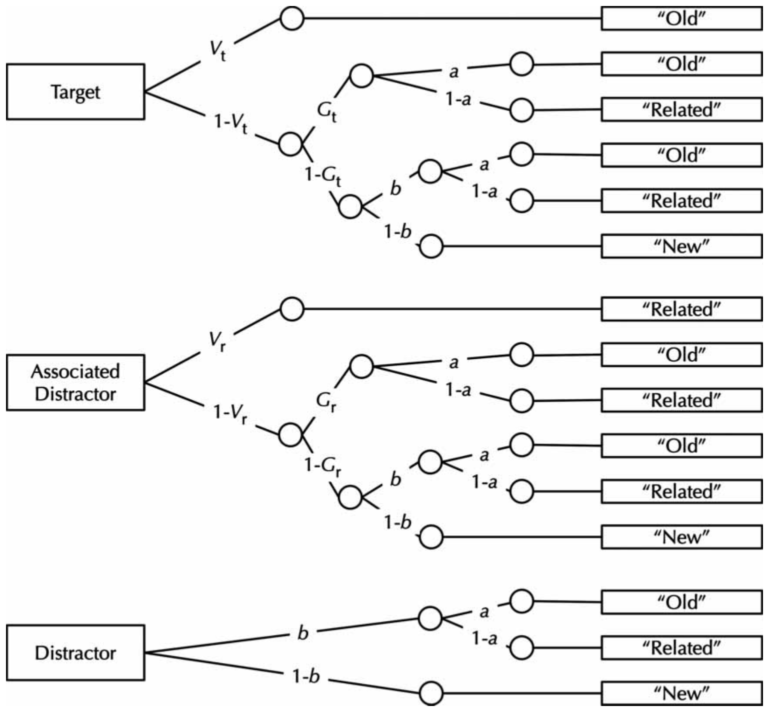

Figure 1. Illustration of the simplified conjoint recognition model of Stahl and Klauer (2008) as adapted for the present purposes. Rectangles on the left side represent the to-be-judged stimuli; rectangles on the right side represent the categories of participants' memory judgements. Letters along the links represent the probabilities with which certain cognitive processes occur: V_t = probability of retrieving a target's verbatim trace given a target; G_t = probability of retrieving a target's gist trace given a target; V_t = probability of retrieving a target's verbatim trace given an associated distractor; G_r = probability of retrieving a target's gist trace given an associated distractor, b = probability of guessing that an item is either a target ("old") or an associated distractor ("related"); a = probability of guessing that an item is a target ("old").

of recall-to-reject, then parameter V_r should be higher in the cued compared to the control condition. The side effect of a general familiarity increase in the cued condition should be reflected in a general increase in parameters a and b which reflect guesses favouring the decision that an item was old. The effect must be general because it is caused by the category questions at test (and, as such, cannot affect verbatim and gist memory traces) and thus should apply to all cued item types equally. In essence, then, an increase of the parameters a and b in the cued compared to the control condition is expected. The suspected side effect of a tendency to respond consistently was not pursued in Experiment 1b because the response categories differed from those of Experiment 1a, and a conditional analysis would be problematic using multinomial modelling due to unequal and, in some cases, small number of cases.

Method

Participants. Participants were 53 adults (mostly students), 41 of whom were female. The mean age of participants was 24 (between 18 and 43) years. All participants were native German speakers and none of them had participated in Experiment 1a.

Materials. The item material as well as study and test list construction were identical to those of Experiment 1a.

Procedure. The procedure was identical to that of Experiment 1a with two exceptions. First, instead of an old/new recognition decision, the use of the simplified conjoint recognition model (Stahl & Klauer, 2008) required participants to identify the type of the test item and to respond with *old*, *related*, or *new*. Participants responded by pressing the *old* (right arrow), *related* (down arrow), or *new* (left arrow) key on a keyboard. Second, asking participants to classify the items required informing them that test lists included associated distractors that were from the same category as a studied word.

Design. The design of Experiment 1b differed from that of Experiment 1a only in that the dependent variable was the response frequency for the recognition decisions. Given a sample size of 53 and $\alpha = .05$, an effect of size $d_z = 0.5$ could be detected with a probability of $1\text{-}\beta = .97$.

Results

Participants responded correctly to 67% of the category questions ($SE = 1.30$) and to 67% of the item questions ($SE = 1.55$). Table 3 displays the proportions of participants' responses (old, related, or new) to the different item types (cued/control targets, cued/control distractors, and cued/control associated distractors).

Behavioural data. The proportion of related responses for cued associated distractors (.63) was higher than for control associated distractors (.53), $t(52) = 3.03$, $p = .002$, $d_z = 0.42$. The proportion of new responses for cued standard distractors (.61) was lower than for control standard distractors (.70), $t(52) = 2.92$, $p = .003$, $d_z = 0.40$. For targets, the proportion of old responses was higher in the cued condition (.80) compared to the control condition (.75), $t(52) = 2.59$, $p = .007$, $d_z = 0.36$.

Concerning the conditional analysis, the response (yes or no) to the unrelated category question preceding control items did not have an effect on the proportion of related responses for associated distractors (.53 vs .50), $t(52) = 0.74$, $p = .461$, $d_z = 0.10$, on the proportion of new responses for standard distractors (.69 vs .71), $t(50) = 0.22$, $p = .824$, $d_z = 0.03$, or on the proportion of old responses for targets (.76 vs .75), $t(50) = 0.26$, $p = .794$, $d_z = 0.04$. The proportion of related responses for cued associated distractors was higher after acceptance of the corresponding category question (.84) than after rejection (.21), $t(46) = 12.85$, $p < .001$, $d_z = 1.89$. For cued standard distractors, the proportion of new responses was lower after acceptance (.17) than after rejection (.90), $t(46) = 17.29$, $p < .001$, $d_z = 2.55$. The proportion of old responses for cued targets was higher after acceptance (.88) than after rejection (.60), $t(47) = 5.07$, $p < .001$, $d_z = 0.74$.

Multinomial modelling. Table 4 displays the parameter estimates and the results of the relevant significance tests. Parameter estimation and significance tests were performed using multiTree (Moshagen, 2010). The hypothesis that category cueing has an effect on the probability of a certain cognitive process can be tested by setting the parameters representing these processes to be equal across cueing conditions. If this constraint leads to a statistically significant decrease in model fit, it can be concluded that the parameter value differs across cueing conditions. Without constraints, the baseline model was saturated and fitted the data well, $G^2 = 0.34$.

The predicted effect of category cueing on recall-to-reject was observed (parameter V_r): As can be seen in Table 4, cueing category exemplars by their corresponding category names increased the probability of recall-to-reject (.00 vs .30), ΔG^2 $(df = 1) = 6.25$, $p = .012$; ΔG^2 is approximately χ^2 distributed under the null hypothesis. However, cueing had no effect on parameter V_t, ΔG^2 $(df = 1) = 1.94$, $p = .164$. As hypothesised, parameter a (.13 vs .25) and parameter b (.30 vs .39) were increased when category exemplars were cued by their corresponding category names, ΔG^2 $(df = 1) = 11.45$, $p = .001$ and ΔG^2 $(df = 1) = 12.11$, $p = .001$, respectively.

Discussion

The results of Experiment 1b strengthen the interpretation of the results of Experiment 1a. Category cueing increases the probability of recall-to-reject. The results also clearly demonstrate that side effects such as a general familiarity increase can occur which may foil the reduction of false alarms to associated distractors.

Although Experiments 1a and 1b provide clear evidence that category cueing increases the probability of recall-to-reject, it would be worthwhile to find a cueing manipulation where the effect of recall-to-reject is not impaired by side effects.

EXPERIMENT 2A

The main goal of Experiment 2a was to avoid the side effects of category cueing in order to achieve a reduction in the absolute level of the false alarm rate. This might be possible by simply using phonological association instead of semantic association as the similarity manipulation. Associated distractors in Experiment 2a 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). The combination of semantic cues and phonological association prevents a direct relation between category question (e.g., animal?) and the subsequent item (e.g., house). Therefore the category question should not lead to an increase in familiarity for the item following the question. As

TABLE 3
Proportion of old, related, and new responses for different item conditions in Experiment 1b and 2b

Experiment	*Item condition*	*Unconditional*			*Conditional*							
					Preceding category accepted			*Preceding category rejected*				
		M (*SE*)	*M* (*SE*)	*M* (*SE*)	*M* (*SE*)	*M* (*SE*)	*M* (*SE*)	*M* (*SE*)	*M* (*SE*)	*M* (*SE*)	*t* (*df*)	d_z
		"old"	"related"	"new"	"old"	"related"	"new"	"old"	"related"	"new"		
1b	Non-cued targets	**.75 (03)**	.12 (.02)	.13 (.02)	**.76 (.03)**	.11 (.03)	.13 (.02)	**.75 (.03)**	.11 (.02)	.14 (.02)	0.26 (50)	0.04
	Non-cued distractors	.04 (.01)	.26 (.03)	**.70 (.03)**	.03 (.01)	.27 (.03)	**.69 (.04)**	.03 (.01)	.26 (.04)	**.71 (.04)**	0.22 (50)	0.03
	Non-cued associated distractors	.08 (.01)	**.53 (.03)**	.38 (.03)	.08 (.02)	**.53 (.03)**	.39 (.03)	.10 (.02)	**.50 (.04)**	.40 (.04)	0.74 (52)	0.10
	Cued targets	**.80 (.02)**	.09 (.02)	.11 (.02)	**.88 (.02)**	.11 (.02)	.01 (.00)	**.60 (.05)**	.08 (.03)	.32 (.05)	5.07 (47) *	0.74
	Cued distractors	.10 (.01)	.30 (.03)	**.61 (.03)**	.16 (.03)	.67 (.04)	.17 (.04)	.05 (.01)	.06 (.02)	.90 (.02)	17.29 (46) *	2.55
	Cued associated distractors	.11 (.02)	**.63 (.03)**	.26 (.03)	.10 (.02)	**.84 (.03)**	.06 (.02)	.12 (.03)	**.21 (.04)**	.67 (.05)	12.85 (46) *	1.89
2b	Non-cued targets	**.77 (.02)**	.06 (.01)	.17 (.02)	.78 (.02)	.05 (.01)	.17 (.02)	.77 (.02)	.06 (.01)	.17 (.02)	0.69 (56)	0.09
	Non-cued distractors	.11 (.01)	.15 (.02)	.74 (.03)	.11 (.02)	.16 (.02)	**.73 (.03)**	.12 (.02)	.16 (.03)	.72 (.03)	0.09 (56)	0.01
	Non-cued associated distractors	.13 (.01)	**.56 (.02)**	.31 (.02)	.14 (.02)	**.58 (.03)**	.28 (.02)	.13 (.02)	**.54 (.03)**	.33 (.03)	1.24 (55)	0.17
	Cued associated distractors	.07 (.01)	**.75 (.02)**	.18 (.02)	.05 (.01)	**.83 (.02)**	.12 (.02)	.09 (.01)	**.62 (.03)**	.29 (.03)	5.56 (56) *	0.74

Standard errors in parentheses. Values printed in bold indicate correct responses. The table shows relative frequencies of responses both in overall terms and depending on the answer to the preceding category question. For the conditional analysis referring to the correct responses (old responses for targets, related responses for associated distractors and new responses for standard distractors), *t*-values, degrees of freedom (*df*) and effect sizes (d_z) are also listed. * $p < .001$

TABLE 4
Model-based results of the simplified conjoint recognition model for Experiments 1b and 2b

Experiment	Parameter	Cueing condition: Control	Cued	$\Delta G^2_{(df=1)}$	p
1b	V_t	.73 (.69, .77)	.77 (.73, .81)	1.94	.164
	G_t	.30 (.19, .42)	.25 (.10, .40)	0.33	.566
	V_r	.00 (.00, .26)	.30 (.16, .43)	6.25	**.012**
	G_r	.45 (.30, .60)	.39 (.25, .53)	0.62	.431
	a	.13 (.09, .18)	.25 (.19, .30)	11.45	**.001**
	b	.30 (.26, .34)	.39 (.36, .43)	12.11	**.001**
2b	V_t	.73 (.71, .76)			
	G_t	.16 (.08, .23)			
	V_r	.38 (.31, .44)	.65 (.60, .69)	70.03	< **.001**
	G_r	.32 (.24, .40)	.29 (.19, .39)	0.27	.603
	a	.42 (.36, .47)			
	b	.26 (.24, .29)			

The table shows parameter estimates (.95 confidence intervals of the parameter estimates in parentheses) and the goodness-of-fit statistic G^2 corresponding to the restriction that parameters do not differ between the control and cued condition. p values smaller than .05 (printed in bold) indicate that the implemented restrictions are not compatible with the data, as a result of which the hypothesis that parameters are identical between cueing conditions must be rejected. V_t = probability of retrieving a target's verbatim trace given a target; G_t = probability of retrieving a target's gist trace given a target; V_r = probability of retrieving a target's verbatim trace given an associated distractor; G_r = probability of retrieving a target's gist trace given an associated distractor, b = probability of guessing that an item is either a target ("old") or an associated distractor ("related"); a = probability of guessing that an item is a target ("old").

an added benefit, changing the similarity manipulation should also eliminate the tendency to respond consistently to the category and item questions. The relation between category question and the subsequent item is now indirect and phonological and 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 subsequent item *house* only becomes apparent if the word *mouse* is recalled. It is exactly this situation in which we expect participants effectively to use recall-to-reject. Therefore we expect the false alarm rate for cued associated distractors to be lower than that for control associated distractors.

With regard to an analysis of the responses depending on the answer to the preceding category question (conditional analysis for short), the predictions made for Experiment 1a again apply. For control items acceptance or rejection of the preceding category question is not expected to have any effect. Due to recall-to-reject, the false alarm rate for cued associated distractors is expected to be lower given that the preceding category question is accepted. If the side effects of semantic category cueing can successfully be avoided, then the conditional results for cued associated distractors should conform to predictions.

Method

Participants. Participants were 64 adults (mostly students), 46 of whom were female. The mean age of participants was 24 (between 19 and 38) years. All participants were native German speakers. None had participated in Experiment 1a or 1b.

Materials. As in Experiments 1a and 1b, participants studied category exemplars and were later tested for recognition memory. Examples of all item types are shown in Table 5. Again, the category questions preceding all control items were unrelated to the subsequent test items. The category question for cued associated distractors referred to the studied exemplar that rhymed with the subsequent test item.[1] Please note that,

[1] There was no cued condition for standard distractors or targets. As standard distractors are not presented themselves, nor are any associated items presented during study, no item exists that could cue a standard distractor. In order to create an indirect cued condition for targets one would have to present rhyme pairs in the study list, which the authors decided against because this should reduce the effectiveness of recall-to-reject (Brainerd et al., 2003; Gallo, 2004). This issue is addressed in the General Discussion.

TABLE 5
Examples for the different item conditions in Experiments 2a and 2b

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

For simplicity, the table does not show items of the study list referring to the category question of control items.

in contrast to Experiments 1a and 1b, a test item was never an exemplar of the preceding category.

Items were taken from two item pools.[2] Cued as well as control associated distractors were taken from the category pairs item pool, which consisted of 40 category pairs; that is, two categories whose exemplars rhymed (i.e., were phonologically similar) and were orthographically similar (e.g., building–house; animal–mouse). 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. Selection of stimuli was further constrained by the need to ensure that only the two category exemplars of one category pair rhyme and that all category exemplars belonged to just 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. These constraints required that the targets and standard distractors as well as the unrelated category questions prior to control items were taken from a second item pool. The second item pool consisted of 140 single category names with one category exemplar each. Categories were randomly allocated to item types and category questions for each participant. 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 appropriate category question. All items of the test list were presented in random order. Each test list was composed of eight control targets, four control distractors, four control associated distractors, and four cued associated distractors.

Procedure. The procedure was identical to that of Experiment 1a 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. As in Experiment 1a, participants were not informed about details of study and test list construction (i.e., that two words which rhyme could not occur during study and that, as in Experiments 1a and 1b, only one exemplar of each category was presented during study), and they were not given explicit recall-to-reject instructions.

Design. Experiment 2a employed a one-factorial design with item type (control target, control distractor, control associated distractor, cued associated distractor) as within-participant variable. The dependent variable was the false alarm rate of the different distractor types.

Given $N = 64$ participants and $\alpha = .05$, it was possible to detect a false recognition effect of size $d_z = 0.94$ (population effect size derived from the sample effect size observed in a pilot study) with a probability of $1\text{-}\beta > .99$. The comparison between the false alarm rate for control and cued associated distractors defined the effect of cueing on recall-to-reject. This comparison is rather conservative given that an increase of the false alarm rate was found for cued associated distractors in the previous experiments despite recall-to-reject (due to side effects).

Results

Participants responded correctly to 67% of the category questions ($SE = 0.97$) and to 85% of the item questions ($SE = 0.96$). The proportions of old responses to the different item types are shown in Table 2.

The false alarm rate for control associated distractors ($M = .15$) was significantly higher than

[2] Although the use of separate item pools may, in principle, confound comparisons, the constraints for the construction of the stimulus material were inevitable to ensure that an adequate number of trials was available for analysis. Furthermore, the use of two item pools seemed justified because all items for the critical comparison (that between cued and control associated distractors) were taken from the same item pool.

for control standard distractors ($M = .11$), 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 ($M = .13$) was significantly lower than that for control associated distractors ($M = .15$), $t(63) = 2.14$, $p = .018$, $d_z = 0.30$. A comparison between cued associated distractors (.13) and control standard distractors (.11) revealed no differences in the false alarm rates, $t(63) = 1.42$, $p = .159$, $d_z = 0.18$.

The conditional analysis revealed that the responses to the preceding unrelated category question had neither an effect on the false alarm rates for standard distractors (.12 vs .10), $t(63) = 0.89$, $p = .375$, $d_z = 0.11$, and associated distractors (.16 vs .16), $t(63) = 0.26$, $p = .795$, $d_z = 0.03$, nor on the hit rates for targets (.80 vs .83), $t(63) = 1.81$, $p = .075$, $d_z = 0.23$. However, the false alarm rate for cued associated distractors was significantly lower if the category question had been accepted ($M = .10$) than if it had been rejected ($M = .19$), $t(63) = 4.04$, $p < .001$, $d_z = 0.51$.

Discussion

As in Experiment 1a, cueing associated distractors with a category question about the corresponding studied item eliminated the false recognition effect. Moreover, in Experiment 2a the false alarm rate for cued associated distractors was reduced significantly and down to the level of that for control standard distractors. This finding indicates that the side effects observed in Experiment 1a were successfully avoided. Furthermore, the results support the interpretation of Experiment 1a in that category cueing increases the probability of recall-to-reject. The results of the conditional analysis provide additional support for these conclusions. 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 because recollecting a category exemplar should lead to the acceptance of the corresponding category question and make recall-to-reject, and hence the rejection of the subsequent associated distractor, very likely.

To summarise, the elimination of the false recognition effect found in Experiment 1a was found again in Experiment 2a, and the false alarm rate was reduced even though no explicit recall-to-reject instructions were given. What is more, the reduction of the false alarm rate and the results of the conditional analysis lead to the conclusion that category cueing did not give rise to side effects (increase of familiarity and tendency to respond consistently) when phonological associations were used.

EXPERIMENT 2B

The simplified conjoint recognition model (Stahl & Klauer, 2008) was used in order to provide additional evidence that category cueing increases the probability of recall-to-reject as interpreted in Experiment 2a. With respect to behavioural data, an increase of recall-to-reject in the cued condition compared to the control condition should be reflected in an increase in related responses for associated distractors. The conditional analysis should not show any differences of response frequencies for all control items after acceptance or rejection of the preceding category question. However, the proportion of related responses for cued associated distractors is expected to be higher after acceptance of the preceding category question due to recall-to-reject.

For multinomial modelling, one set of the three model trees depicted in Figure 1 for the control condition and one additional tree for cued associated distractors were needed. The three response categories of the tree for cued associated distractors allow the computation of two additional parameters (i.e., V_r and G_r) in addition to the six parameters of the control condition. If category cueing does increase the probability of recall-to-reject, then parameter V_r should be higher in the cued compared to the control condition. However, the conclusion that category cueing does not give rise to side effects that oppose recall-to-reject when phonological association is used, cannot be tested in Experiment 2b as the present design does not allow for the computation of separate parameters a and b for cued and control items.

Method

Participants. Participants were 57 adults (mostly students), 40 of whom were female. The mean age of participants was 24 (between 16 and 38) years. All participants were native German speakers. None of them had participated in Experiments 1a, 1b, or 2a.

Materials. The items as well as study and test list constructions were identical to those of Experiment 2a.

Procedure. The procedure of Experiment 2b was identical to that of Experiment 2a with two exceptions. First, participants were asked to identify the type of test item and to respond with *old*, *related*, or *new*. Second, participants were informed that test lists included associated distractors that rhymed with a studied word.

Design. The design of Experiment 2b differed from that of Experiment 2a only in that the dependent variable was the response frequency for the recognition decisions. Given a sample size of 57 and $\alpha = .05$, an effect of size $d_z = 0.5$ could be detected with a probability of $1\text{-}\beta = .98$.

Results

Participants responded correctly to 66% of the category questions ($SE = 1.19$) and to 72% of the item questions ($SE = 1.30$). The lower part of Table 3 displays the frequencies of participants' responses (old, related, or new) to the different types of category exemplars (control target, control distractor, control associated distractor, cued associated distractor).

Behavioural data. The proportion of related responses for cued associated distractors (.75) was higher than for control associated distractors (.56), $t(56) = 10.00$, $p < .001$, $d_z = 1.34$. Concerning the conditional analysis, the response (yes or no) to the unrelated category question preceding control items did not have an effect on the proportion of related responses for associated distractors (.58 vs .54), $t(55) = 1.24$, $p = .220$, $d_z = 0.17$, on the proportion of new responses for standard distractors (.73 vs .72), $t(56) = 0.09$, $p = .926$, $d_z = 0.01$, or on the proportion of old responses for targets (.78 vs .77), $t(56) = 0.69$, $p = .494$, $d_z = 0.09$. The proportion of related responses for cued associated distractors was higher after acceptance of the corresponding category question (.83) than after rejection (.62), $t(56) = 5.56$, $p < .001$, $d_z = 0.74$.

Multinomial modelling. The saturated baseline model fitted the data very well, $G^2 = 0.00$. The lower part of Table 4 displays the parameter estimates and the results of the relevant significance tests. The probability of recall-to-reject (parameter V_r) was higher for cued (.65) than for control associated distractors (.38), ΔG^2 $(df = 1) = 70.03$, $p < .001$.

Discussion

The results of Experiment 2b strengthen the interpretation of the results of Experiment 2a. The recognition data of Experiment 2a and 2b as well as the results obtained using the multinomial model in Experiment 2b suggest that the probability of recall-to-reject can be increased substantially by category cueing. Moreover, as the probability of parameter V_r was above zero by a considerable margin in the control condition, the results of Experiment 2b demonstrate that recall-to-reject also plays an important role in single item recognition without category cueing.

GENERAL DISCUSSION

The present series of experiments yielded a consistent pattern of results. Category cueing had the expected effects on false recognition data as indirect indicators of recall-to-reject (all experiments) as well as on model parameters representing recall-to-reject directly (Experiments 1b and 2b). In Experiment 1a category cueing eliminated the false recognition effect observed for semantically associated distractors. Moreover, in Experiment 2a a reduction in the absolute level of the false alarm rate was found for phonologically associated distractors. These results demonstrate that cueing category exemplars by their corresponding category names increased the probability of recall-to-reject. This interpretation is strengthened by the behavioural and modelling results obtained using the simplified conjoint recognition model in Experiments 1b and 2b. Given that no explicit recall-to-reject instructions were given, we can conclude that individuals are capable of identifying the study and test list structures themselves (in line with e.g., Hall & Kozloff, 1970) and spontaneously use recall-to-reject without being instructed to (in line with e.g., Gallo, 2004; Leding & Lampinen, 2009). Thus, recall-to-reject seems to be a memory-editing mechanism relevant to everyday memory. However, more research has to be done to understand what conditions have to be met for recall-to-reject to occur spontaneously.

Furthermore, the results of the present experiments provide further evidence that recall-to-reject also plays an important role in single item recognition tasks in which associated distractors have not been studied. When associated distractors are used that have been studied, like in plurality discrimination tasks (where only singular/plural is changed) or in other discrimination tasks, in which associated distractors were studied but in a different context, be it in a different list (e.g., Rotello & Heit, 2000), in a different modality (e.g., Light, LaVoie, Valencia-Laver, Albertson Owens, & Mead, 1992), or in a different format (e.g., Gallo et al., 2007), the need to rely on recall and consequently the probability of recall-to-reject should be greater. The recall of studied items should be particularly important in associative recognition where associated distractors are composed of a mixture of features from studied items. For example, in a typical associative recognition task, associated distractors are recombinations of words from studied word pairs (e.g., Rotello & Heit, 2000). Similarly, associated distractors in the memory conjunction paradigm are compound words with old but recombined parts (e.g., blackbird when blackmail and jailbird were studied); this paradigm has become increasingly popular especially for the study of recall-to-reject (Arndt & Jones, 2008; Jones, 2006; Jones & Atchley, 2006; Lampinen et al., 2004; Odegard et al., 2005). According to the differences in recognition tasks, evidence for recall-to-reject has been more frequent in discrimination or associative recognition tasks compared to single item recognition using unstudied associated distractors (also see Malmberg, 2008; Rotello & Heit, 2000).

Furthermore, the probability of recall-to-reject should increase with the tendency of the associated distractor to trigger recalling the corresponding studied item (Odegard et al., 2005). Given that associated distractors in an associative recognition task are composed of a mixture of features from studied items, they should be better retrieval cues for their corresponding studied targets than associated distractors in a standard item recognition task. These differences between tasks could explain why more evidence for recall-to-reject has often been found in associative compared to item recognition tasks (e.g., Rotello & Heit, 2000). Category names seem to be good retrieval cues for category exemplars, as is evident from the fact that recall-to-reject was very effective in the present experiments. With regard to recognition data, the false recognition effect in Experiment 1a was not only reduced but eliminated, and in Experiment 2a an absolute reduction of false recognition for associated distractors was achieved. With regard to the model-based results, category cueing increased the probability of recall-to-reject by .30 in Experiment 1b and by .27 in Experiment 2b.

However, the structure of the study list may have played an important role in this effectiveness. As associated distractors were similar to only one item of the study list, recalling the corresponding learned 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; two in Rotello & Heit, 1999), 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* of those presented items. Thus 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). Hence the usefulness of category cues in reducing the false recognition effect should decrease when the number of similar items presented at study increases. This manipulation might be used in the future to investigate further how much of a role recall-to-reject plays in recognition memory when it is not facilitated.

The present results provide evidence for the hypothesis that category cueing increases the probability of recall-to-reject and can reduce false recognition. However, the results also demonstrated that side effects can occur which oppose the reduction of false alarms. The higher level of false alarms to cued standard distractors in Experiment 1a and the lower proportion of new responses to cued standard distractors as well as the increase in parameters *a* and *b* in Experiment 1b was indicative of an increase in familiarity. The simultaneous familiarity increase can explain why evidence for recall-to-reject is rather rare in terms of old/new recognition data. For example, a ''null effect'' of repetition has been observed with respect to raw false alarm rates of associated distractors (Jones & Jacoby, 2001; Lampinen et al., 2004; Tussing & Greene, 1999). Raw recognition data will only be indicative of recall-to-reject if the effect of recall-to-reject is stronger than the effect of familiarity, or if data are compared to a control condition that is solely influenced by familiarity. However, although raw recognition data cannot provide a pure measure of recall-to-reject, they are important for several reasons. First, only recognition data can reveal if manipulations that increase recall-to-reject could be used to reduce or even eliminate false memories in the context of eyewitness testimony and forensic interviews (Brainerd & Reyna, 2002b). Second, whereas many methods such as self-reports and the simplified conjoint recognition model require to inform participants at least to some extent about the experimental paradigm, the old/new recognition tests used in the present study reveal that individuals are capable of using recall-to-reject spontaneously.

A second side effect of the cueing procedure observed in Experiment 1a was that participants tended to respond such that their response to a particular item question was consistent with their response to the preceding category question. As was obvious from the conditional analysis in Experiment 1a 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. In Experiment 1b the probability of correctly rejecting standard distractors was lower when the preceding category question had been accepted. Given that no exemplar of the category had been presented during study, the incorrect *yes* response to the category question as well as the incorrect responses to the item question must be due to guessing. The tendency to be consistent when guessing becomes obvious in the conditional analysis, but influences the unconditional data too. However, it is difficult to make precise predictions about the exact effect on the unconditional data. It seems possible that the tendency to respond consistently works in opposition to recall-to-reject. For example, the tendency to respond consistently should increase the false alarms for associated distractors in Experiment 1a because the probability of accepting the preceding category question will be higher than rejecting it (because an exemplar of the category has been learned) and the initial probability of rejecting an associated distractor should be higher than accepting it. In contrast to the familiarity increase, the tendency to respond consistently is a side effect that should be restricted to the specific design used in the present work. Furthermore, the results of Experiments 2a and 2b indicate that side effects could be successfully avoided by 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 for different stimulus materials such as pictures. Memory representations for pictures are usually more distinctive than for words (e.g., Israel & Schacter, 1997). Thus, using category names as retrieval cues might be more effective and possible side effects might not occur even when semantic association is used.

In summary, category cueing increases the probability of recall-to-reject and has the potential to reduce false recognition errors. The results demonstrate that recognition is influenced by both familiarity and recall-like processes, consistent with the assumptions of dual-process theories. In addition, category cueing can have side effects that oppose recall-to-reject. Furthermore, the present work provides further evidence that participants spontaneously use recall-to-reject without explicitly being instructed to do so, and that recall-to-reject does also play an important role in single item recognition.

Manuscript received 11 November 2009
Manuscript accepted 2 June 2010

REFERENCES

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.

Arndt, J., & Jones, T. C. (2008). Elaborative processing and conjunction errors in recognition memory. *Memory & Cognition*, *36*, 899–912.

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

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

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., & Howe, M. L. (2009). Trichotomous processes in early memory development, aging, and neurocognitive impairment: A unified theory. *Psychological Review*, *116*, 783–832.

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., & Payne, D. G. (2002). Dual-retrieval processes in free and associative recall. *Journal of Memory and Language*, *46*, 120–152.
Coane, J. H., & McBride, D. M. (2006). The role of test structure in creating false memories. *Memory and Cognition*, *34*, 1026–1036.
Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, *82*, 407–428.
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.
Epstein, M. L., Dupree, D. A., & Gronikowski, L. A. (1979). Encoding specificity and contextual similarity. *Bulletin of the Psychonomic Society*, *14*, 177–180.
Erdfelder, E., Auer, T.-S., Hilbig, B. E., Aßfalg, A., Moshagen, M., & Nadarevic, L. (2009). Multinomial processing tree models. A review of the literature. *Zeitschrift für Psychologie/Journal of Psychology*, *217*, 108–124.
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. (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.
Hall, J. W., & Kozloff, E. E. (1970). False recognitions as a function of number of presentations. *American Journal of Psychology*, *83*, 272–279.
Higham, P. A., & Vokey, J. R. (2004). Illusory recollection and dual-process models of recognition memory. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, *57A*, 714–744.
Hintzman, D. L., & Curran, T. (1994). Retrieval dynamics of recognition and frequency judgements: Evidence for separate processes of familiarity and recall. *Journal of Memory and Language*, *33*, 1–18.
Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, *6*, 65–70.
Hudson, R. L., & Austin, J. B. (1970). Effect of context and category name on the recall of categorized word lists. *Journal of Experimental Psychology*, *86*, 43–47.
Israel, L., & Schacter, D. L. (1997). Pictorial encoding reduces false recognition of semantic associates. *Psychonomic Bulletin & Review*, *4*, 577–581.
Jones, T. C. (2005). Study repetition and the rejection of conjunction lures. *Memory*, *13*, 499–515.
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., Brown, A. S., & Atchley, P. (2007). Feature and conjunction effects in recognition memory: Toward specifying familiarity for compound words. *Memory and Cognition*, *35*, 984–998.
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.
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.
Leding, J. K., & Lampinen, J. M. (2009). Memory conjunction errors: The effects of presentation duration and study repetition. *Memory*, *17*, 597–607.
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.
Malmberg, K. J. (2008). Recognition memory: A review of the critical findings and an integrated theory for relating them. *Cognitive Psychology*, *57*, 335–384.
Mandler, G. (2008). Familiarity breeds attempts: A critical review of dual-process theories of recognition. *Perspectives on Psychological Science*, *3*, 390–399.
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.

Moshagen, M. (2010). multiTree: A computer program for the analysis of multinomial processing tree models. *Behavior Research Methods*, *42*, 42–54.
Odegard, T. N., Koen, J. D., & Gama, J. M. (2008). Process demands of rejection mechanisms of recognition memory. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *34*, 1296–1304.
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.
Ratcliff, R., & McKoon, G. (1988). A retrieval theory of priming in memory. *Psychological Review*, *95*, 385–408.
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. (2001). Recall processes in recognition memory. In D. L. Medin (Ed.), *The psychology of learning and motivation: Advances in research and theory* (pp. 183–221). San Diego, CA: 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.
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.
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. (2008). A simplified conjoint recognition paradigm for the measurement of gist and verbatim memory. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *34*, 570–586.
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.
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.
Wortschatz Lexikon Deutsch. (2005). Universität Leipzig. Retrieved 9 April 2005, from http://wortschatz.informatik.uni-leipzig.de
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.