Journal of Memory and Language 68 (2013) 333–349

Contents lists available at SciVerse ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

Rethinking familiarity: Remember/Know judgments in free recall

Laura Mickes [a], Travis M. Seale-Carlisle [b], John T. Wixted [b,*]

[a] Department of Psychology, University of Warwick, Coventry CV4 7AL, UK
[b] Department of Psychology, University of California, San Diego, United States

ARTICLE INFO

Article history:
Received 27 April 2012
revision received 22 December 2012
Available online 28 February 2013

Keywords:
Recollection
Familiarity
Recall
Recognition

ABSTRACT

Although frequently used with recognition, a few studies have used the Remember/Know procedure with free recall. In each case, participants gave Know judgments to a significant number of recalled items (items that were presumably not remembered on the basis of familiarity). What do these Know judgments mean? We investigated this issue using a source memory/free-recall procedure. For each word that was recalled, participants were asked to (a) make a confidence rating on a 5-point scale, (b) make a Remember/Know judgment, and (c) recollect a source detail. The large majority of both Remember judgments and Know judgments were made with high confidence and high accuracy, but source memory was nevertheless higher for Remember judgments than for Know judgments. These source memory results correspond to what is found using recognition, and they raise the possibility that Know judgments in free recall identify the cue-dependent retrieval of item-only information from an episodic memory search set. In agreement with this idea, we also found that the temporal dynamics of free recall were similar for high-confidence Remember and high-confidence Know judgments (as if both judgments reflected retrieval from the same search set). If Know judgments in free recall do in fact reflect the episodic retrieval of item-only information, it seems reasonable to suppose that the same might be true of high-confidence Know judgments in recognition. If so, then a longstanding debate about the role of the hippocampus in recollection and familiarity may have a natural resolution.

Published by Elsevier Inc.

Introduction

Tulving (1972) drew a theoretical distinction between episodic memories and semantic memories. Although both are cue-dependent, episodic memories are linked to time and space and involve self-referential information (e.g., remembering the day of your college graduation), whereas semantic memories are not linked to time and space and do not involve self-referential information (e.g., remembering the number of bones in the human body). Later, Tulving (1985) developed a metacognitive technique known as the Remember/Know procedure that was designed to identify episodic and semantic memories. As originally conceived, a "remember" judgment was intended to identify retrieval from episodic memory and a "know" judgment was intended to identify retrieval from semantic memory. Thus, for example, you may *remember* that you threw your black graduation cap at the end of the morning graduation ceremony, but you may simply *know* that there are 206 bones in the human body. The Remember/Know procedure was not wedded to a particular method of testing memory, so Tulving (1985) used it with free recall, cued recall and recognition.

Not long after it was created to distinguish between episodic and semantic memory, others started using the Remember/Know procedure to distinguish between recollection and familiarity. According to dual process theories, recognition decisions are based either on the recollection

* Corresponding author. Address: Department of Psychology, 0109, University of California, San Diego, La Jolla, CA 92093, United States.
E-mail address: jwixted@ucsd.edu (J.T. Wixted).

0749-596X/$ - see front matter Published by Elsevier Inc.
http://dx.doi.org/10.1016/j.jml.2013.01.001

of contextual details or on the familiarity of the item itself (see Mandler, 2008; Yonelinas, 2002, for reviews). The use of the Remember/Know procedure to investigate recollection and familiarity quickly caught on, and it is now used for that purpose vastly more often than it is used to study episodic and semantic memory. In the present article, we consider the possibility that Tulving's (1985) intended use of the procedure was somewhat closer to the mark than its current usage. More specifically, we present evidence suggesting that Know judgments reflect the cue-dependent retrieval of item-only information (though from episodic memory, not semantic memory). According to this view, Remember judgments reflect the retrieval of item plus associative information from an episodic search set, whereas Know judgments reflect the retrieval of item plus little or no associative information from that same episodic search set.

The meaning of a Remember judgment is essentially the same regardless of whether it is used to identify retrieval from episodic memory, as in its original usage, or to identify the occurrence of recollection during a test of recognition memory, which is how it is typically used today. According to Tulving (1985), to *remember* an episode is to appreciate that it was part of one's personal past. Such an appreciation necessarily involves retrieving contextual detail about the encoded episode, which is what Remember judgments are used to identify in studies of recognition memory. However, the meaning of a Know judgment differs considerably depending on whether it is used to identify retrieval from semantic memory or to identify the occurrence of context-free familiarity. In Tulving's (1985) original conceptualization, to *know* a fact is to be aware that the information was successfully retrieved from memory despite the absence of any accompanying sense of personal involvement or any accompanying contextual detail. A fact can be known in this sense whether it was retrieved by an act of recall (e.g., recalling the number of bones in the human body) or by an act of recognition (e.g., knowing that the word "judo" appeared on a list). By contrast, in studies of recognition memory framed by dual-process theory, to know that an item appeared on a list means having the experience of familiarity in the absence of recollection. Such an experience is usually thought to be specific to recognition because, on a recall test, no item is presented to generate a familiarity signal (Quamme, Yonelinas, Kroll, Sauve, & Knight, 2004; Wixted & Squire, 2010; but see Brainerd & Reyna, 2010).

In the present research, we investigated Remember/Know judgments in free recall. If free recall is based on recollection, as is typically assumed, and if Know judgments reflect familiarity-based memories, then free recall should be exclusively characterized by Remember judgments. In that case, the Remember/Know procedure would not be very useful in free recall, which may explain why the Remember/Know procedure is almost always used in conjunction with recognition. For example, in 2010 alone, at least 39 articles published results from the Remember/Know procedure used with recognition, but we are aware of only four studies that have used it with free recall since it was developed in 1985 (Hamilton & Rajaram, 2003; McCabe, Roediger, & Karpicke, 2010; McDermott, 2006; Tulving, 1985). Somewhat surprisingly, however, all four of these studies found that a substantial fraction of freely recalled words (e.g., 20%) received know judgments. What do these know judgments mean, and what do they suggest, if anything, about the meaning of familiarity?

Tulving (1985) found that as the quality of the retrieval cue increased from free recall to cued recall over successive retrieval attempts, the proportion of items recalled increased (as might be expected), but the proportion of retrieved items that received Remember judgments decreased. Correspondingly, the proportion of retrieved items that received Know judgments increased. From this, Tulving (1985) hypothesized that the use of autonoetic consciousness is a function of the retrieval support provided. When retrieval support is low (as in free recall), episodic trace information must be high for successful retrieval to occur, but when it does occur, it is accompanied by autonoetic consciousness (and by Remember judgments). When retrieval support is high (as in cued recall), the information can be successfully retrieved from semantic memory, in which case it is accompanied by noetic awareness (and by Know judgments). However, using a between subjects design, Hamilton and Rajaram (2003) found that as retrieval support increased and overall performance improved, the proportion of retrieved items that received Remember responses remained equivalent, suggesting that more complete retrieval cues can also facilitate retrieval from episodic memory. Although both Tulving (1985) and Hamilton and Rajaram (2003) measured the effect of retrieval support on Remember judgments, neither study investigated the meaning of Know judgments in free recall.

McCabe et al. (2010) conducted the only previous study that was specifically designed to identify the meaning of Know judgments in free recall. They argued that Know judgments reflect automatic memory, a concept that is related to the notion of familiarity. Early interpretations of familiarity were based on notions of perceptual integration (Mandler, 1980) or perceptual fluency (Jacoby & Dallas, 1981) associated with a test item. These perception-based conceptualizations would appear to limit the experience of familiarity to recognition. However, the occurrence of Know judgments in free recall suggested to McCabe et al. (2010) that a version of dual-process theory may have applicability to free recall as well. The same possibility was briefly mentioned by McDermott (2006) in a study of the testing effect on free recall, which found that successful recall can be associated with both Remember and Know judgments.

The dual-process account offered by McCabe et al. (2010) holds that Remember judgments reflect consciously controlled retrieval, whereas Know judgments reflect automatic memory (i.e., memory that occurs without an act of volition). Automatic memory is thought to give rise to the subjective experience of familiarity on recognition tests, but it is not necessarily a perception-based account and thus could apply to free recall as well. Using a variant of Jacoby's (1991) process-dissociation procedure adapted to free recall, McCabe et al. (2010) found that dividing attention at study reduced the recollection estimate but did not affect the automatic estimate. Correspondingly, in a second experiment, they found that dividing attention at study re-

duced the number of Remember judgments but had no effect on the number of Know judgments. In light of these findings, they argued that automatic memory can occur in free recall as well as in recognition.

Dissociations between Remember and Know judgments like the one reported by McCabe et al. (2010) have often been reported in the recognition memory literature over the years, but the meaning of those dissociations is not clear because they can usually be accommodated by a single-process signal-detection model (Dunn, 2004; Wixted & Stretch, 2004). The main complicating issue is that Remember judgments are typically made with high confidence and high accuracy, whereas Know judgments are typically made with lower confidence and lower accuracy. Thus, any difference in the pattern of results associated with Remember and Know judgments – a difference that is ordinarily attributed to a differential effect of an experimental manipulation on different memory processes – can be just as easily explained by a differential effect of the experimental manipulation on strong and weak memories arising from a single process.

To address the strength confound that characterizes most studies that use the Remember/Know procedure, Wixted and Mickes (2010) investigated strong familiarity-based recognition memories. They found that old/new accuracy for high-confidence Know judgments was often similar to old/new accuracy for high-confidence Remember judgments (i.e., these Remember and Know judgments were equally strong). Even so, source recollection was much higher for high-confidence Remember judgments than for high-confidence Know judgments, suggesting that participants can report accurately on the content of their memories even after controlling for old/new confidence and accuracy (see also Ingram, Mickes, & Wixted, 2012).

Here, we investigate similar issues in the context of free recall. We ask, for example, whether Know judgments in free recall are associated with lower confidence and lower accuracy than Remember judgments, as is typically true in recognition. In addition, we ask whether Remember and Know judgments in free recall reflect different degrees of source memory even after they are equated for confidence and accuracy. We set out to address these issues in four experiments. In Experiment 1, Remember/Know judgments and confidence ratings were collected in a free recall task involving lists of words drawn from different semantic categories (as in past research on this topic). In agreement with past research, we found that a substantial fraction of responses were, indeed, accompanied by Know judgments. Beyond that, and unlike in recognition, we found that the large majority of Know judgments in free recall were made with high confidence and high accuracy. In Experiment 2, we provided criterial source information at study to measure qualitative differences between Remember and Know judgments at retrieval. As in recognition, we found that Know judgments made with high confidence and high accuracy were nevertheless associated with lower source recollection accuracy than high-confidence Remember judgments (attesting to the metacognitive validity of these Know judgments). In Experiment 3, we added a forced recall condition to investigate the possibility that participants were using a generate-recognize strategy, in which case Know judgments in free recall might reflect familiarity-based decisions after all. The results weighed against that possibility and in favor of the idea that these Know judgments reflect item-only retrieval from an episodic memory search set. Finally, in Experiment 4, we investigated whether high-confidence Know judgments would be observed even when the study items consisted of a list of unrelated words. Even in that case, a substantial number of recalled words were associated with Know judgments made with high confidence and high accuracy.

Experiment 1

Experiment 1 was a straightforward free recall task, modeled on past research, in which Remember/Know judgments and confidence ratings were provided for each item recalled. Based on previous work (Hamilton & Rajaram, 2003; McCabe et al., 2010; McDermott, 2006; Tulving, 1985), we expected that participants would provide Know judgments to some of the words they recalled, but we also collected confidence ratings on a 5-point scale so we could compare the distribution of confidence ratings for Remember and Know judgments. In addition, we compared accuracy scores to see if Remember judgments mainly reflect correct responses whereas Know judgments reflect a mixture of correct and incorrect responses (as might be expected based on the typically lower accuracy scores for Know judgments in recognition).

Method

Participants

Fifty University of California, San Diego (UCSD) undergraduates participated in exchange for course credit.

Materials

Three lists of 24 words were created by randomly drawing words from six different semantic categories (4 words per category). No two lists used the same semantic categories. The words were selected from category norms (Van Overschelde, Rawson, & Dunlosky, 2004), and we used categories that had at least eight words. Stimuli were presented and responses were recorded with the E-prime program (www.pstnet.com; Psychology Software Tools).

Procedure

Participants signed a consent form, listened to instructions, and were then presented with three lists that contained 24 words each (totaling 72 words). Each word appeared in the center of the screen in black courier font size 18 for 4 s with an inter-stimulus interval of 250 ms. After a 15-s distractor task (an adapted serial sevens test), they were prompted to type as many words as they could recall in any order. After they typed a word, they indicated their confidence that the word appeared during the presentation phase using a 5-point scale (ranging from 1 = low confidence to 5 = 100% sure the word was on the list). Next, they indicated whether their response was based on recollection or familiarity by making a "remember" (by pressing the "r" key) or "know" (by pressing the

"k" key) judgment, respectively. After making a Remember/Know judgment, they were prompted to recall the next word. If they could not recall any more words, they pressed the enter key (once for each remaining word to be recalled) and the next list was presented. Lists 2 and 3 proceeded in the same manner.

The Remember/Know instructions were based on Gardiner and Richardson-Klavehn (2000), and the instructions emphasized that a Remember judgment should be made if anything about the presentation of the word was recalled (such as thoughts the word elicited when it was studied), whereas a Know judgment should be made if the memory of the item was accompanied by no contextual/source detail. Prior to the experiment, participants were given a short practice list and test to acquaint them with the rating scale, and to ensure that they understood the Remember/Know distinction. Any questions that arose after the practice trial were answered before list 1 was presented. Each participant was tested individually in a quiet room.

Results and discussion

Table 1 shows the total number of words recalled following each list (correct plus incorrect) in the first column. The remaining columns show the total number of words recalled broken down in two different ways, first according to whether the recalled words were correct or incorrect and, second, according to whether they were associated with Remember judgments or Know judgments. An analysis of variance performed on the overall number of words recalled showed no significant effects (or trends) across the three lists, and the same was true of an analysis performed on the other measures shown in Table 1 (for all experiments, an alpha level of .05 was used unless otherwise noted). Because performance was similar across lists, the remaining analyses are based on performance aggregated across the 3 lists (72 words in all).

Fig. 1 shows the average number of words recalled (including the incorrect words) across the 3 lists as a function of confidence, separately for Remember and Know judgments. Of the highest confidence responses (i.e., those items that received a rating of 5), 63% received a Remember judgment and 37% received a Know judgment. Of the Remember judgments that were made across all 5 levels of confidence, 86% were made with the highest level of confidence, and of the Know judgments that were made across all 5 levels of confidence, 78% were made with the highest level of confidence. Thus, both Remember and Know judgments were mainly made with high confidence.

Table 1
Overall number of words recalled per list in Experiment 1 (Recalled), and the overall number of words recalled partitioned in two ways, first according to whether they were correct (Correct) or incorrect (Incorrect) and second according to whether they were associated with Remember judgments (Remember) or Know judgments (Know).

List	Recalled	Correct	Incorrect	Remember	Know
1	17.9 (3.1)	16.9 (3.2)	1.0 (1.5)	11.4 (6.1)	6.5 (4.9)
2	17.4 (3.8)	16.4 (3.8)	1.0 (1.4)	10.3 (6.2)	7.1 (6.0)
3	17.6 (3.2)	16.8 (3.4)	0.8 (1.2)	10.7 (7.0)	6.9 (6.3)

Note: Parenthetical values are standard deviations.

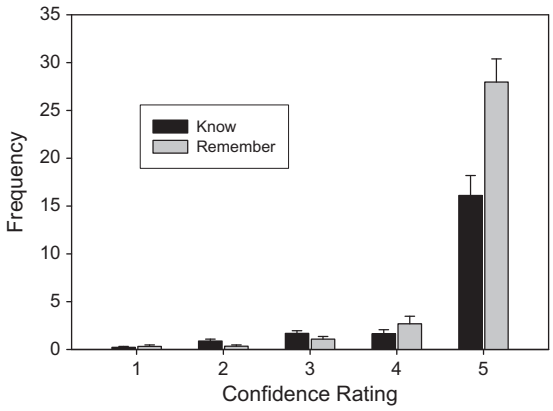

Fig. 1. Average frequency of Remember/Know judgments as a function of confidence rating in Experiment 1 (error bars represent standard errors).

Of the 50 participants in this experiment, 48 made at least 1 Remember judgment and 1 Know judgment. For these 48 participants, the overall accuracy (collapsed across levels of confidence) of Remember judgments (.95, $SD = 0.10$) was significantly higher than the overall accuracy of Know judgments (.88, $SD = 0.20$), $t(47) = 2.37$. Similarly, the average confidence associated with Remember judgments (4.75) was significantly higher than the average confidence associated with Know judgments (4.31), $t(47) = 3.64$. Thus, according to these measures (i.e., according to both confidence and accuracy), Remember judgments reflect stronger memories than Know judgments, as is invariably true of recognition as well (see Wixted & Mickes, 2010). Most participants (41) made at least one high-confidence Remember judgment and one high-confidence Know judgment. For them, the accuracy scores were 0.98 ($SD = .04$) and 0.97 ($SD = 0.06$), respectively. Thus, on average, Know judgments were made with lower confidence and lower accuracy than Remember judgments, but high-confidence Know judgments (which constitute the large majority of Know judgments) were made with very high accuracy.

These results answer the main questions that motivated Experiment 1: Know judgments in free recall occur frequently (in agreement with past research), they do not mainly reflect low-confidence (instead, they mainly reflect high confidence), and high-confidence Know judgments do not reflect low accuracy (instead, like high-confidence Remember judgments, they reflect high accuracy).

Given that Know judgments in free recall do not simply reflect guessing (or weak memory, at least when made with high confidence), the question arises as to what they do reflect. Experiment 2 was designed to shed light on that question by including an explicit source memory attribute at study (memory for which was tested at retrieval).

Experiment 2

In Experiment 2, as each item was presented for study, participants were asked to make an animacy judgment or a size judgment. Later, when a word from the list was re-

called, they were asked to recollect which question accompanied the item at study (in addition to making a confidence rating and a Remember/Know judgment).

Method

Participants

Sixty UCSD undergraduates participated for psychology course credit.

Materials

These were the same as Experiment 1.

Procedure

The procedure was the same as Experiment 1 except that we included a source memory test. During the study phase, each word appeared with one of two questions that required an animacy judgment ("is this item animate or inanimate?") or size judgment ("is this item bigger than a shoebox?"). Immediately following the study phase, participants were tested. After the first list, the testing procedure was the same as in Experiment 1: participants first typed a word, then indicated their confidence that the word was presented, and then made a Remember or Know judgment. The source memory test was a surprise and was administered after the participant had completed recalling as many words from List 1 as possible. The test consisted of presenting each word that was recalled and asking whether that word was associated with the animacy question or the size question at study. For Lists 2 and 3, the procedure was similar, except that the source memory question for each word appeared immediately after both the confidence rating and the Remember/Know judgment were made (for these lists, the source memory test would not be a surprise, so there was no reason to delay it).

Results

Table 2 shows the number of words recalled following each list broken down in the same manner as in Table 1. Unlike in Experiment 1, the number of words recalled following List 1 was noticeably higher than the number of words recalled following both List 2 and List 3 (which were similar to each other). An analysis of variance performed on these data revealed a significant effect of list number for both the number of recalled words, $F(2,118) = 23.29$, and the number of correctly recalled words, $F(2,118) = 19.86$. This effect presumably reflects the fact that, once participants realized that source memory would be tested (following the surprise source test administered after the first list), they devoted extra effort attempting to memorize the source question associated with each word. This extra effort to memorize source details apparently came at the expense of recalling more of the words from the list. Consistent with this explanation, source accuracy following List 1 ($M = 0.58$) was lower than that following both List 2 and List 3 ($M = 0.67$ and 0.64, respectively), whereas the source accuracy scores for Lists 2 and 3 were similar to each other. An analysis of variance performed on these data revealed that the effect of list number on source accuracy was significant, $F(2,118) = 7.13$. When tested separately, the source accuracy score for each of the 3 lists was significantly greater than chance, $t(59) = 4.60$, 8.58 and 5.27, for lists 1, 2 and 3, respectively.

Fig. 2 shows the average number of Remember and Know judgments (both correct and incorrect) for the 3 lists combined as a function of confidence. Most of the recalled words (82%) were recalled with a confidence rating of 5. Of those, 62% were Remember judgments and 38% were Know judgments. Of all Remember judgments, 88% were made with high confidence, and of all Know judgments, 75% were made with high confidence. Thus, as in Experiment 1, the large majority of both Remember and Know judgments were made with high confidence.

Across the 3 lists, the average accuracy of Remember judgments (0.94) significantly exceeded the average accuracy of Know judgments (0.84), $t(58) = 2.79$. In addition, the average confidence associated with Remember judgments (4.77) significantly exceeded the average confidence associated with Know judgments (4.01), $t(58) = 4.99$. Thus, as with Experiment 1 (and as with recognition memory experiments), confidence and accuracy measures indicate that Remember judgments reflect stronger memories than Know judgments.

Of particular interest were the Remember and Know judgments made with high confidence. Not all participants made both high-confidence Remember judgments and high-confidence Know judgments, but the remaining analyses focus on those who did. Forty-six participants (77%) made at least one high-confidence Remember judgment and one high-confidence Know judgment across the 3 lists. The average recall accuracy scores for these Remember and Know judgments were 0.95 ($SD = 0.08$) and 0.91 ($SD = 0.21$), respectively, values that did not differ significantly. The accuracy score for the high-confidence Know judgments was somewhat variable because 5 participants made only one such response. To increase the precision of this measure, we also conducted a separate analysis on participants who made a larger number of responses in each category. For the 30 participants who made at least

Table 2
Overall number of words recalled per list in Experiment 2 (Recalled), and the overall number of words recalled partitioned in two ways, first according to whether they were correct (Correct) or incorrect (Incorrect) and second according to whether they were associated with Remember judgments (Remember) or Know judgments (Know). Source accuracy for each list is also shown.

List	Recalled	Correct	Incorrect	Remember	Know	Source accuracy
1	16.2 (3.3)	15.0 (3.3)	1.2 (1.6)	9.8 (6.0)	6.5 (5.5)	0.58 (0.13)
2	13.3 (4.2)	12.6 (4.1)	0.7 (1.0)	7.6 (5.1)	5.7 (5.1)	0.67 (0.15)
3	12.9 (4.0)	11.9 (3.8)	1.0 (1.6)	7.3 (5.0)	5.6 (5.3)	0.64 (0.21)

Note: Parenthetical values are standard deviations.

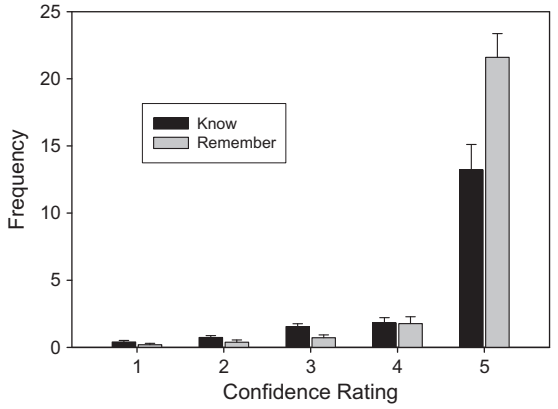

Fig. 2. Average frequency of Remember/Know judgments as a function of confidence rating in Experiment 2 (error bars represent standard errors).

5 high-confidence Remember judgments and 5 high-confidence Know judgments (50%), the recall accuracy scores were 0.95 (SD = 0.07) and 0.96 (SD = 0.06), respectively. Thus, even though, on average, Remember judgments are made with higher confidence and higher accuracy than Know judgments, high-confidence Remember judgments and high-confidence Know judgments were once again highly (and comparably) accurate.

The main question of interest in this experiment concerned source accuracy associated with correct high-confidence Remember and Know judgments. As shown in Fig. 3, for the 45 participants (75%) who made at least 1 correct high-confidence Remember judgment and 1 correct high-confidence Know judgment collapsed across the 3 lists, source accuracy was significantly higher for Remember judgments (M = .68) than for Know judgments (M = .58), $t(44)$ = 2.42. Although lower than that associated with Remember judgments, the source accuracy score for Know judgments was nevertheless significantly greater than chance, $t(44)$ = 2.07, as is typically true of recognition as well (Wixted & Mickes, 2010). Source accuracy remained significantly higher for high-confidence Remember judgments (compared to high-confidence Know judgments) when we performed the analysis on the 30 participants (50%) who made at least 5 correct high-confidence Know judgments (M = .57) and 5 correct high-confidence Remember judgments (M = .71), $t(29)$ = 3.91. Thus, in free recall, Remember and Know judgments would appear to provide valid information about memory for source detail. Once again, however, the source accuracy score for Know judgments was significantly greater than chance, $t(29)$ = 2.33. This is typically true of recognition data as well and it is consistent with the idea that Know judgments are made with respect to a decision criterion placed on a continuous source recollection signal such that Remember judgments are made when *enough* recollection occurs (i.e., not in response to the categorical occurrence of recollection but in response to the memory signal exceeding a criterion). When the source recollection signal is relatively weak (i.e., when it is not strong enough to exceed a decision criterion), a Know judgment is made (cf. Wixted & Mickes, 2010).

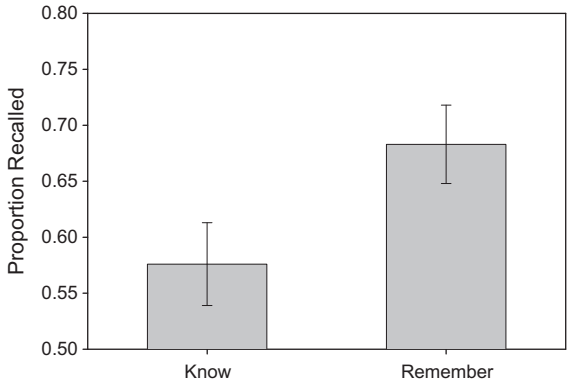

Fig. 3. Source accuracy scores (and associated standard errors) of the 45 participants who made at least one correct high-confidence Remember judgment and one correct high-confidence Know judgment in Experiment 2.

We also analyzed the results from List 1 separately. For that list, the source recall test was unexpected, and the test was administered only after all of the confidence ratings and Remember and Know judgments had been made. Thirty-one participants made at least 1 high-confidence Remember judgment and 1 high-confidence Know judgment on List 1. For these participants, the source accuracy score for Remember judgments (M = 0.62) was significantly greater than chance, $t(30)$ = 2.36, but the source accuracy score for Know judgments (M = .56) was not. However, the source accuracy scores for high-confidence Remember and Know judgments did not differ significantly from each other. Ten participants (17%) made at least 5 high-confidence Remember judgments and 5 high-confidence Know judgments on List 1, and their source accuracy scores did differ significantly (M = .53 for Know judgments, and M = .73 for Remember judgments), $t(9)$ = 3.30. Once again, for these 10 participants, the source accuracy score for Remember judgments was significantly greater than chance, $t(9)$ = 4.20, but the source accuracy score for Know judgments was not. Thus, although these results from List 1 are based on relatively few participants, the data provide some evidence that Remember and Know judgments provide valid information about the content of memory even when the source memory test is unexpected.

Overall, these results indicate that Know judgments in free recall, when made with high-confidence (as the large majority are), are very accurate. Even so, they are associated with significantly less source accuracy than high-confidence Remember judgments. This pattern is very similar to what has been observed in recognition memory experiments (Wixted & Mickes, 2010). One difference is that although high-confidence Know judgments are observed in recognition, they often constitute the minority of Know judgments when memory is tested that way. In free recall, by contrast, the large majority of Know judgments are made with high confidence and high accuracy.

These results attest to the validity of Remember/Know judgments in free recall, and they weigh against the idea that these judgments reflect nothing more than a difference in memory strength (strong vs. weak). Still, they do

not rule out a strength-based interpretation. Although source accuracy differed when confidence and accuracy for item recall were equated at a high level, it is theoretically possible that an undetectable difference in item memory strength (stronger for Remember than Know) remained. Still, the results do show that source recollection differs in the expected direction (greater for Remember than Know) when item memory is associated with high confidence and almost perfect accuracy.

On the surface, the existence of high-confidence Know judgments in free recall seems hard to reconcile with the notion that Know judgments reflect familiarity (again, because no test item is presented to generate a familiarity signal). However, a familiarity-based interpretation would be viable if, during recall, some words are first generated and are then recognized on the basis of familiarity. The generate/recognize explanation is plausible because we used categorized lists of words (as all prior studies of Remember/Know judgments in free recall have done). Thus, for example, a participant might remember that various professions were presented on the list and might use that knowledge to generate a word like "plumber," which might then be recognized solely on the basis of familiarity. In that case, Know judgments in free recall would reflect familiarity-based decisions after all, just as they are often thought to do in recognition. In the next experiment, we used a forced-recall procedure to investigate the possibility that Know judgments in free recall arise from a generate-recognize strategy (e.g., Kintsch, 1970).

Experiment 3

In some ways, the results of Experiments 1 and 2 already weigh against a generate-recognize interpretation of Know judgments in free recall. For example, if generated items were recognized on the basis of familiarity, then we should have observed a wider distribution of confidence ratings associated with Know judgments (as generally occurs in recognition). That is, some generated items would presumably have low familiarity (yielding a confidence rating of 1), others would have somewhat greater familiarity (yielding a confidence rating of 2), and so on. Instead, we found that Know judgments were characterized by an almost discontinuous distribution, with the large majority receiving the highest level of confidence (a pattern that also applied to Remember judgments).

Nevertheless, it is possible that, for whatever reason, participants typically used a high confidence criterion on the familiarity scale before electing to type a word during the recall test. A high decision criterion would account for the relative absence of familiarity-based decisions made with lower confidence (the kind of Know decisions that often predominate on recognition tests). If participants did use a high criterion for making familiarity-based decisions, it stands to reason that they also covertly generated additional items from the list associated with lower degrees of familiarity (items that were not overtly recalled despite being covertly generated).

Experiment 3 was designed to investigate whether such items were in fact covertly generated by adding a forced recall condition (following Roediger & Payne, 1985). In a forced-recall procedure, participants who have studied a list of *n* items on a list are asked to recall *n* items, even if they have to guess. The hypothetical data in Fig. 4 illustrate a pattern of results that would be consistent with the generate-recognize hypothesis. The high-confidence Remember and Know judgments would look much like those found in Experiment 1 and 2, but because participants are also forced to recall additional items, they would now be expected to overtly produce additional covertly generated items associated with lower levels of familiarity and, therefore, lower levels of confidence (including additional correct items that appeared on the list). That is, the predicted pattern would consist of a selective increase in correct (and incorrect) Know judgments made with confidence ratings of 1 through 4. Unlike Know judgments, Remember judgments should be essentially unaffected.

By contrast, if high-confidence Know judgments reflect item-only recollection instead of familiarity, a different pattern should be obtained. Specifically, because there are no additional low-familiarity items to output, participants should produce many more low-confidence guesses (few if any of which are correct).

In Experiment 3, we also measured reaction times (RTs) associated with Remember and Know judgments because the dynamics of free recall help to differentiate between an automatic-memory interpretation of Know judgments in free recall (McCabe et al., 2010) and an item-only-recollection interpretation. That is, it seems reasonable to predict that automatic responses (which receive a Know judgment) would come to mind faster than the consciously-controlled responses (which receive a Remember judgment). Indeed, the notion that automatic memory is faster than memory arising from consciously controlled search is widely accepted (Yonelinas, 2002). However, if Remember and Know judgments both reflect the outcome of consciously controlled retrieval from the same episodic memory search set, then standard models of free recall predict that their retrieval dynamics should instead be

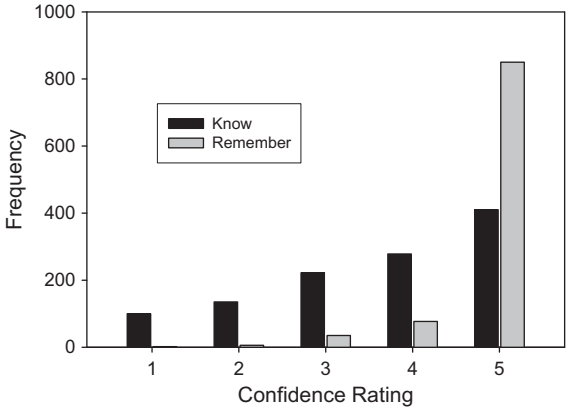

Fig. 4. Hypothetical data predicted by the generate–recognize account. If the generate–recognize strategy were employed, then correct Remember judgments would not be affected, but correct Know judgments would be. As shown, Know judgments would show a selective increase for correctly recalled words.

governed by their respective memory strengths (Wixted, Ghadisha, & Vera, 1997).

According to relative strength models of free recall, of which SAM (search of associative memory) is the best known example (Gillund & Shiffrin, 1984), the probability that an item will be sampled from a search set is a function of its strength relative to the summed strength of the other items in the set. When strong and weak items are intermixed in the same search set, the strong items will be recalled more quickly than the weak items (Wixted et al., 1997). Thus, if words associated with Remember and Know judgments are retrieved from the same search set, then Remember judgments should be associated with *faster* (not slower) responding than Know judgments. This prediction is based on the fact that memories associated with Remember judgments are stronger than those associated with Know judgments according to measures of confidence and accuracy. However, this account further predicts that the speed of recall should be approximately the same for Remember and Know judgments once their strengths are approximately equated (i.e., when Remember and Know judgments involve comparably high levels of confidence and accuracy).

Method

Participants

Sixty UCSD students were randomly assigned to a group (30 in the free recall group and 30 in the forced recall group).

Materials

These were the same as Experiments 1 and 2.

Procedure

The procedure was similar to Experiment 1, except we added a forced recall condition. The participants who were assigned to the forced recall condition were instructed to type 24 words after each 24-item list, even if they had to guess. They were also reminded that if they did make a random guess, they should use a "1" on the confidence rating scale.

For both the free and forced recall conditions, we also recorded the time required to recall each word. The reaction times (RTs) were measured with respect to a prompt that appeared at the beginning of the recall period and that appeared again immediately after the confidence rating and Remember/Know judgment had been entered for the previously recalled word. Timing began with the presentation of each recall prompt and continued until the word was entered (indicated by pressing the enter key). Thus, the recorded data consist of a series of RTs, each of which reflects the time required to search for and then type a word.

Results and discussion

Table 3 shows the number of words recalled following each list broken down in the same manner as in Tables 1 and 2. The maximum possible number of words recalled per list (correct plus incorrect) was 24, so, as shown in the first column of the table, participants in the forced-recall group largely followed our instructions to produce 24 words per list. An analysis of variance performed on the overall number recalled revealed a main effect of condition (free vs. forced), $F(1,58) = 112.3$, as expected, but neither the effect of list number nor the interaction between condition and list number were significant. An analysis of variance performed on the data for correctly recalled words revealed a marginally significant effect of condition (free vs. forced), $F(1,58) = 3.75$, but, once again, neither the effect of list number nor the interaction between condition and list number were significant. Compared to participants in the free recall group, participants in the forced recall group produced approximately 6 more words per list during the recall period and, of those, approximately 1.6 were correct, on average.

The frequency distributions of correct responses for the free recall group (Fig. 5A) and the forced recall group (Fig. 5B) show that, as in the previous experiments, the large majority of correct responses were made with the highest level of confidence (for both Remember and Know judgments). Moreover, the slightly higher number of items correctly recalled in the forced recall group compared to the free recall group was also concentrated at the high end of the confidence scale and consisted of an increase in high-confidence Remember judgments (coupled with a smaller decrease in high-confidence Know judgments). The frequency distributions of *incorrect* responses for the free (Fig. 5C) and forced recall (Fig. 5D) groups show that, as might be expected, there were many more incorrect responses for the forced recall group. Most of these were concentrated at the low end of the confidence scale, but there were also more Remember errors in the forced recall group than the free recall group at the high confidence end of the scale. These high-confidence errors mainly reflect repetitions of words that had been correctly recalled earlier in the recall period.

As indicated earlier, if participants were covertly generating correct items with relatively low familiarity but electing not to produce them (because the level of familiarity would not support a confidence rating of 5), then when those words are produced in the forced recall condition, they should be associated with Know judgments made with confidence ratings of 1 through 4. Contrary to this prediction, forced recall resulted in an increase in the number of *Remember* judgments made with confidence ratings of 5, without any increase in the number of Know judgments (or Remember judgments) made with lower confidence. The average number of correct Know judgments made with a confidence rating of 1 through 4 was 1.0 ($SD = 1.03$) in the free recall condition and 1.3 ($SD = 1.11$) in the forced recall condition (a difference that did not approach statistical significance). This is the key test of the generate-recognize account, which predicts an increase in the number of Know judgments made with less-than-high confidence in the forced recall condition. The average number of correct Remember judgments made with a confidence rating of 1 through 4 was 1.1 ($SD = 1.82$) in the free recall condition and 1.01 ($SD = 1.57$) in the forced recall condition (a difference that also did not approach statistical significance). For responses made with high confidence (a rating of 5), the number of Know judgments

Table 3
For the free recall (Free) and forced recall (Forced) conditions of Experiment 3, overall number of words recalled per list (Recalled), and the overall number of words recalled partitioned in two ways, first according to whether they were correct (Correct) or incorrect (Incorrect) and second according to whether they were associated with Remember judgments (Remember) or Know judgments (Know).

Condition	List	Recalled	Correct	Incorrect	Remember	Know
Free	1	17.6 (4.2)	16.2 (4.0)	1.4 (1.8)	11.4 (6.6)	6.2 (5.7)
	2	17.7 (3.9)	16.5 (3.9)	1.3 (1.3)	12.1 (6.4)	5.6 (6.0)
	3	16.9 (4.2)	15.8 (4.0)	1.0 (1.2)	11.8 (5.6)	5.1 (5.4)
Forced	1	23.5 (0.9)	18.3 (3.3)	5.2 (3.5)	15.8 (6.3)	7.6 (6.2)
	2	23.7 (1.0)	18.0 (3.7)	5.7 (3.8)	17.0 (5.0)	6.7 (4.9)
	3	23.7 (0.8)	17.0 (4.4)	6.7 (4.4)	16.9 (6.3)	6.8 (6.3)

Note: Parenthetical values are standard deviations.

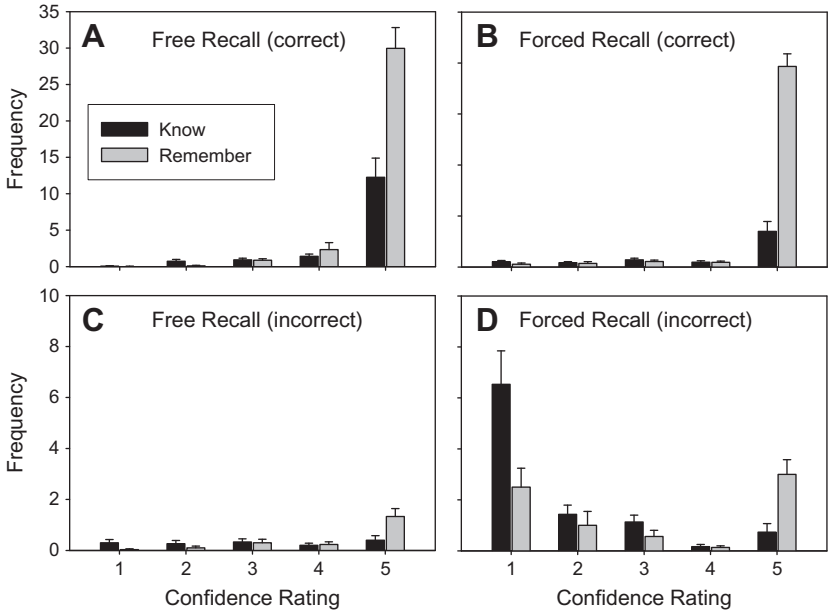

Fig. 5. Average frequency of Remember/Know judgments as a function of confidence rating in Experiment 3 for correctly recalled words (Panel A: Free Recall Group; Panel B: Forced Recall Group) and for incorrectly recalled words (Panel C: Free Recall Group; Panel D: Forced Recall Group).

showed a non-significant *decrease* in the forced recall condition while the number of Remember judgments showed a significant increase. More specifically, the average number of correct Know judgments made with a confidence rating of 5 was 4.1 ($SD = 4.80$) in the free recall group and 2.3 ($SD = 3.60$) in the forced recall group (a difference that did not approach statistical significance). The average number of correct Remember judgments made with a confidence rating of 5 was 9.9 ($SD = 5.17$) in the free recall group and 13.1 ($SD = 4.55$) in the forced recall group, a difference that was significant, $t(58) = 2.49$.

Why were slightly more words recalled with high confidence in the forced-recall condition compared to the free-recall condition, and why did the difference between the two conditions exhibit opposite trends for Remember and Know judgments made with high confidence? The time spent recalling words before giving up was under the control of the participant, and (as might be expected) it differed for the free and forced recall conditions. A participant's cumulative "search time" for a given list was defined as the time from the presentation of the prompt to recall words from the list (presented at the beginning of the recall period) to the moment when the enter key was pressed for the last correctly recalled word in that recall period (with the clock stopped during the time required to make confidence ratings and Remember/Know judgments for each recalled word). An analysis of variance performed on cumulative search time scores revealed a main effect of list number, $F(2,116) = 8.06$), reflecting the fact that participants spent less time searching as list number increased (94.2 s, 76.4 s, and 70.2 s for lists 1, 2 and 3, respectively), and a main effect of group, $F(1,58) = 6.07$, reflecting the fact that participants in the forced recall group searched longer ($M = 88.7$ s, $SD = 30.3$ s) than participants in the free recall group ($M = 71.9$ s, $SD = 22.1$ s). It is well known that, in free recall, participants tend to stop searching before they have recalled all of the items they are capable of recalling and that more items will be recalled if the search continues longer (e.g., Roediger & Thorpe, 1978). Moreover, it makes sense that participants in the forced recall condition would continue searching for a longer period of time given that they needed to pro-

duce a total of 24 words. Thus, the slightly higher number of correctly recalled words in the forced recall group likely reflects the extra time they spent searching (yielding additional words recalled with high confidence), not the overt production of covertly generated items that were recognized with lower levels of confidence on the basis of familiarity (as a generate-recognize account would predict).

In addition to searching longer (and correctly recalling slightly more words as a result), participants in the forced-recall condition were apparently more inclined to declare that successfully recalled words were remembered. That is, these participants apparently used a lower criterion to say "remember." Although it is not clear why such a change in bias would occur, it would explain why Remember judgments increased and Know judgments decreased in the forced recall condition compared to the free recall condition.

Whatever the explanation for the increased number of high-confidence Remember judgments in the forced-recall condition, the key point is that, overall, the observed pattern of results shows no hint that participants were relying on a generate-recognize strategy. Had that been the case, we should have observed an increase in the number of correctly recalled words associated with Know judgments made with confidence ratings of less than 5 in the forced recall condition.

Confidence and accuracy of remember and know judgments

As in the previous experiments, Remember judgments were indicative of stronger memory (measured by confidence and accuracy) than Know judgments, but the difference in accuracy was no longer apparent when the analysis was restricted to words recalled with high confidence. Of the 30 participants in the free recall condition, 29 made at least 1 Remember judgment and 1 Know judgment (including all recalled words regardless of confidence). The average confidence rating for Remember judgments ($M = 4.75$, $SD = 0.48$) was significantly higher than the average confidence rating for Know judgments ($M = 3.96$, $SD = 0.95$), $t(28) = 3.73$. Similarly, the average accuracy (i.e., correct divided by correct plus incorrect) for Remember judgments ($M = 0.94$, $SD = 0.06$) was significantly higher than the average accuracy for Know judgments ($M = 0.81$, $SD = 0.27$), $t(27) = 2.68$. However, the accuracy difference was no longer apparent when the analysis was restricted to words recalled with the highest confidence rating of 5. For the 21 participants who made both a high-confidence Remember judgment and a high-confidence Know judgment in the free recall condition, the average accuracy of these Remember and Know judgments were 0.96 ($SD = 0.05$) and 0.98 ($SD = 0.04$), respectively. Thus, on average, Remember judgments indicate stronger memories than Know judgments, but the difference is eliminated when the analysis is limited to words recalled with high confidence. Once again, it is important to note that an unmeasurable difference in strength between high-confidence Remember and Know judgments might still exist (because the dependent measures are at the top of their respective scales), but it seems clear that both Remember and Know judgments reflect strong memories when the analysis is limited to words recalled with high confidence.

The same strength patterns were observed in the forced recall condition. Once again, 29 participants made at least 1 Remember judgment and 1 Know judgment. The average confidence rating for Remember judgments ($M = 4.56$, $SD = 0.56$) was significantly higher than the average confidence rating for Know judgments ($M = 2.76$, $SD = 1.21$), $t(28) = 7.29$. Similarly, the accuracy of Remember judgments ($M = 0.86$, $SD = 0.11$) was significantly higher than the accuracy for Know judgments ($M = 0.49$, $SD = 0.29$), $t(28) = 7.21$. When the analysis was limited to words recalled with high confidence, 20 participants made at least 1 Remember judgment and 1 Know judgment. The average accuracy of these Remember and Know judgments were 0.93 ($SD = 0.07$) and 0.85 ($SD = 0.30$), respectively, a difference that did not approach significance. The large standard deviation for the Know judgment accuracy score reflects the fact that two participants made only a single know judgment, and both were incorrect (so their accuracy scores were 0). Excluding those two participants, the average accuracy of the high-confidence Remember and Know judgments were 0.93 ($SD = 0.07$) and 0.94 ($SD = 0.09$), respectively.

Remember/Know reaction time analyses

The results presented thus far suggest that Know judgments in free recall do not reflect familiarity-based decisions arising from a generate-recognize process, but it might be argued that they nevertheless reflect some other form of automatic memory. It is widely assumed that an automatic process occurs faster than a consciously controlled process (Yonelinas, 2002). If so, and if Know judgments reflect an automatic process, then Know judgments in free recall should be made faster than Remember judgments. By contrast, if Remember and Know judgments reflect retrieval from the same episodic memory search set, and if recall follows a relative strength rule (as is widely assumed), then the opposite result should be observed. The reason is that, as just described, Remember judgments were associated with stronger memories than Know judgments. This relative strength account further predicts that if Remember and Know judgments are equated for strength at a high level (as they appear to be when the analysis is limited to items recalled with high confidence), no difference will be observed in the speed of recall.

To test these predictions, we analyzed reaction times (RTs) measured from the onset of the prompt to recall a word to the moment the enter key was pressed (after the word was typed in). The prompt was presented at the start of recall period and again each time a Remember/Know judgment was entered for a word that had just been recalled. In the free recall condition, the mean RT for correct Remember judgments ($M = 4.81$ s, $SD = 4.71$ s) was much faster than the mean RT for correct Know judgments ($M = 10.08$ s, $SD = 14.37$ s), a difference that was marginally significant, $t(27) = 1.86$, $p = .07$. However, the Know RTs for two participants were extreme outliers (one had an RT of 61.3 s and the other 54.2 s) and the Remember RT for another subject was also an extreme outlier (29.0 s). Each of these scores was more than 3 standard deviations from their respective means, and, in a visual plot of the data, they stood out conspicuously from the remaining distribu-

tion of RT scores. With those three outliers excluded, the mean RT for Know judgments (M = 6.41 s, SD = 5.10 s) was still longer than the mean RT for Remember judgments (M = 3.97 s, SD = 1.13 s), and the difference was significant, $t(24) = 2.71$. This finding is the opposite of what would be predicted by an automatic memory interpretation of Know judgments, but it is consistent with a relative strength model of free recall (because Know judgments are associated with lower confidence and lower accuracy than Remember judgments).

The relative strength model predicts that the RT difference will be eliminated once the strength difference between Remember and Know judgments is eliminated (as it appeared to be when the analysis was limited to words recalled with high confidence). For the 21 participants in the free recall condition who made both a high-confidence Remember judgment and a high-confidence Know judgment, the mean RT for correct Remember judgments (M = 3.79 s, SD = 1.10 s) was somewhat faster than the mean RT for correct Know judgments (M = 6.08 s, SD = 10.46 s), but the difference was not significant. However, the Know RT for one subject was an extreme outlier (the RT for that subject was 51.76 s, which was more than 4 standard deviations above the mean). With that one outlier excluded, the mean RT for high-confidence Know judgments (M = 3.90 s, SD = 2.34 s) was very similar to that for high-confidence Remember judgments (M = 3.79 s, SD = 1.10 s). Thus, when Remember and Know judgments were essentially equated for confidence and accuracy – that is, they were equated for strength at a high level – the RTs were equated as well (cf. Rotello & Zeng, 2008). This finding is again consistent with a relative strength model of free recall in which both Remember and Know judgments reflect retrieval from the same episodic memory search set.

Similar results were observed in the forced recall condition. Of the 28 participants who made at least 1 correct Remember judgment and 1 correct Know judgment, the mean RT for Remember judgments (M = 4.74 s, SD = 2.09 s) was significantly faster than the mean RT for Know judgments (M = 8.69 s, SD = 5.09 s). However, when the strength difference was minimized by limiting the analysis to words recalled with high confidence, the RT difference disappeared. For the 18 participants who made at least 1 correct high-confidence Remember judgment and 1 correct high-confidence Know judgment, the mean RT for correct Remember judgments (M = 4.02 s, SD = 1.17 s) was virtually identical to the mean RT for correct Know judgments (M = 4.15 s, SD = 1.79 s).

All of these results are consistent with a relative strength model of free recall according to which Remember and Know judgments reflect retrieval from the same episodic memory search set, but they seem hard to reconcile with the view that Know judgments in free recall reflect automatic memory.

Experiment 4

Previous studies of Remember/Know judgments in free recall, including the first 3 experiments reported here, have followed Tulving's (1985) lead of using categorized lists. However, it seems natural to wonder if the phenomenon of interest (namely, high-confidence Know judgments that are made with high accuracy) is limited to categorized lists. In Experiment 4, we tested memory for lists of unrelated words. As in Experiment 1, participants were asked to make a confidence rating and a Remember/Know judgment for each word that was recalled. Also, as in Experiment 3, we measured RTs associated with Remember/Know judgment.

Method

Participants

Thirty UCSD students participated for psychology course credit.

Materials

Words that ranged in length from 3 to 8 letters, and ranged in concreteness (from moderate to high; 450–700), were pulled from the MRC Psycholinguist Database (Coltheart, 1981). That search yielded a large pool of words (1816) of which, 72 words were randomly selected to make up three lists of 24. Each participant studied the same 72 words, but words varied in their list location and presentation order varied for each participant.

Procedure

The procedure was the same as Experiment 1 except that words on the study list were presented for 5 s each.

Results and discussion

Table 4 shows the number of words recalled broken down in the same manner as in Tables 1–3. An analysis of variance performed on these data showed no significant effects (or trends) across the three lists for any of the dependent measures. Thus, the remaining analyses are based on performance aggregated across the 3 lists. Fig. 6 shows the average number of words recalled (including the incorrect words) across the 3 lists as a function of confidence, separately for Remember and Know judgments. Once again, the large majority of words received the highest rating of confidence (5), and, as with the related lists used in Experiments 1–3, a substantial proportion of those words (27%) received Know judgments.

Of the 30 participants in this experiment, 29 made at least 1 Remember judgment and 29 made at least 1 Know

Table 4
Overall number of words recalled per list in Experiment 4 (Recalled), and the overall number of words recalled partitioned in two ways, first according to whether they were correct (Correct) or incorrect (Incorrect) and second according to whether they were associated with Remember judgments (Remember) or Know judgments (Know).

List	Recalled	Correct	Incorrect	Remember	Know
1	14.2 (3.9)	13.2 (3.6)	1.0 (1.5)	9.8 (4.4)	5.4 (4.1)
2	14.8 (4.7)	13.8 (4.7)	1.0 (1.4)	10.1 (5.1)	5.8 (4.6)
3	14.7 (4.4)	14.1 (4.2)	0.6 (0.9)	11.0 (5.0)	5.2 (4.8)

Note: Parenthetical values are standard deviations.

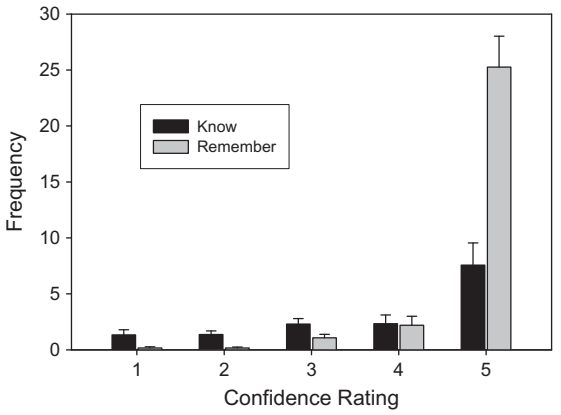

Fig. 6. Average frequency of Remember/Know judgments as a function of confidence rating in Experiment 4 (error bars represent standard errors).

judgment (28 made at least 1 Remember and 1 Know judgment). As in the previous experiments, the overall accuracy of Remember judgments (.98, $SD = 0.05$) was significantly higher than the overall accuracy of Know judgments (.79, $SD = 0.26$), $t(27) = 4.06$. Similarly, the average confidence associated with Remember judgments (4.78, $SD = 0.34$) was significantly higher than the average confidence associated with Know judgments (3.78, $SD = 1.04$), $t(27) = 5.57$. Thus, as with the categorized lists used in Experiments 1, 2 and 3 (and as is typically true in recognition memory experiments), Remember judgments reflect stronger memories than Know judgments. Most participants (21) made at least one high-confidence Remember judgment and one high-confidence Know judgment. For them the accuracy scores were 0.98 ($SD = .05$) and 0.93 ($SD = 0.13$), respectively, a difference that was marginally significant, $t(20) = 2.00$, $p = .059$. Thus, although high-confidence Know judgments in free recall are quite accurate, they do not appear to be quite as accurate as high-confidence Remember judgments (unlike in the 3 previous experiments).

The RT data for correct Remember and Know judgments reflect the strength measures presented above. Collapsed across confidence ratings, the average RT for correct Remember judgments (4.84 s, $SD = 1.86$ s) was significantly faster than the average RT for correct Know judgments (9.11 s, $SD = 6.00$ s), $t(26) = 3.50$. When the same analysis was performed on the 21 participants who made at least 1 high-confidence Remember judgment and 1 high-confidence Know judgment, the RTs were more similar but still showed a marginally significant difference. The average RT for correct Remember judgments was (4.38 s, $SD = 1.86$ s), and the average RT for correct Know judgments (6.46 s, $SD = 4.50$ s), $t(20) = 1.96$, $p = .064$.

These data reinforce the conclusions from the previous experiments. More specifically, the fact that a substantial number of high-confidence Know judgments (made with high accuracy) occur even under the conditions used in Experiment 4 lends further credence to the notion that they do not reflect the product of a generate-recognize process (which seems much less likely to play a role in the free recall of unrelated words). In addition, the fact that Know judgments were again made more slowly than Remember judgments weighs against an automatic memory interpretation of Know judgments. Instead, the data are consistent with a relative strength model that assumes that both Remember and Know judgments reflect retrieval from the same episodic memory search set. Because they reflect stronger memory (according to confidence and accuracy measures), words associated with Remember judgments should be (and are) retrieved more quickly than words associated with Know judgments. When strength was nearly equated (by analyzing words recalled with high confidence), Remember and Know RTs were nearly equated as well.

General discussion

The experiments reported here replicated prior work showing that participants use both Remember and Know judgments in free recall, just as they do in recognition (Hamilton & Rajaram, 2003; McCabe et al., 2010; McDermott, 2006; Tulving, 1985). Beyond that, we also found that (1) the large majority of Know judgments were made to words that were recalled with high confidence (as was also true of Remember judgments), (2) recall accuracy (i.e., the probability that the recalled word appeared on the list) was very high for both high-confidence Remember and high-confidence Know judgments, (3) source accuracy was significantly lower for high-confidence Know judgments compared to high-confidence Remember judgments (attesting to the validity of these judgments), (4) Know judgments in free recall appear not to reflect familiarity-based decisions arising from a generate-recognize strategy, and (5) reaction times associated with Remember and Know judgments correspond to a relative strength rule and are consistent with the idea that both may reflect consciously controlled retrieval from the same episodic memory search set. These results have implications for the understanding of free recall, but they may also have implications for recognition as well, particularly with respect to a longstanding debate in the cognitive neuroscience literature about the role of the hippocampus in recollection and familiarity. We consider first the implications for recall.

What do know judgments in free recall mean?

Tulving (1985) argued that know judgments, whether they are made during recall or recognition, reflect item-only information retrieved from semantic memory. Elsewhere, he also argued that "Access to, or actualization of, information in the episodic system tends to be deliberate and usually requires conscious effort, whereas in the semantic system it tends to be automatic" (Tulving, 1983, p. 46). Here, we propose a different idea. Our interpretation holds that both Remember and Know judgments in free recall reflect cue-dependent, consciously controlled retrieval from a single episodic memory search set. This interpretation agrees with Tulving's (1985) account in one respect (namely, that context-free, item-only recall can occur) but not in another respect (namely, that item-only recall on a free recall test reflects semantic memory). Basically, we advance an interpretation that is more consistent with

Tulving's (1972) original view, which held that free recall for a recently presented list of items is a test of episodic memory, whereas semantic memory instead reflects the recall of information learned across multiple encoding episodes (e.g., knowledge of the number of bones in the human body). According to this view, we did not test retrieval from semantic memory in our experiments.

If this interpretation is correct, then both Remember and Know judgments would reflect the outcome of a consciously controlled search process, and the difference between them would be in the amount of source information that is retrieved when an item is recovered from the search set. That is, the difference between Remember and Know judgments in free recall would not be that one judgment reflects recollection and the other familiarity, or that one reflects a consciously controlled retrieval process and the other an automatic process. Instead, the difference would be that one reflects the consciously controlled retrieval of item-plus-source information from an episodic memory search set and the other reflects the consciously controlled retrieval of item information (with limited or no source information) from that same search set.

Our episodic memory interpretation is illustrated in Fig. 7. This figure depicts a model in which the presentation of 24 items on a list creates a search set of 24 memory traces. During retrieval, these traces are sampled by a random search process with replacement. This is a simplified version of the recall process envisioned by SAM (Gillund & Shiffrin, 1984), which assumes that recall is based on a relative strength rule. If all of the items in the search set have the same memory strength (i.e., in a pure-strength list), then they would all have an equal likelihood of being sampled throughout the recall period. This equal-strength version of the model predicts that cumulative recall (i.e., the number of items recalled as a function of time spent recalling in the recall period) should be characterized by a negatively accelerated exponential growth to asymptote. The model further predicts that if some items in the search set are stronger than others (e.g., in a mixed-strength list), then the strong items would be preferentially sampled throughout the recall period (thereby delaying the recall of the weaker items). In that case, the strong items would be characterized by a faster rate of approach to asymptote than the weak items. Obviously, a more realistic model would take into account the semantic relationships between different subsets of categorized words, would include a stopping rule, and might also include provisions for output interference. However, these added complexities are not needed to illustrate the basic idea, which has often been used to characterize the dynamics of free recall (e.g., Rohrer & Wixted, 1994; Unsworth, 2007; Unsworth, Spillers, & Brewer, 2012; Wixted et al., 1997).

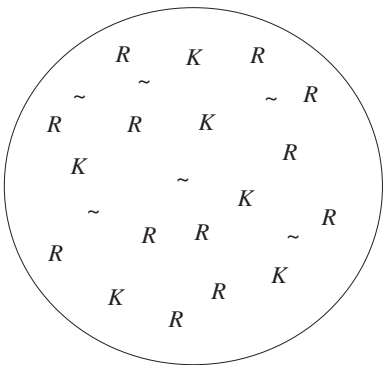

Fig. 7. Hypothetical episodic memory search set created by presenting 24 items on a study list. The ~ symbol represents a non-recoverable item, K represents a recoverable item that is associated with little or no source information, and R represents a recoverable item associated with more substantial source information.

Of the 24 traces shown in Fig. 7,6 are such that, when sampled, they are not sufficiently intact to recover the corresponding word (represented by the symbol "~"). As in SAM, these traces take time to sample, but they result in no response (either overt or covert). Of the remaining 18 traces, 6 are such that they contain mainly item information (represented by a "*K*" in Fig. 7). When these traces are sampled, the originally presented word is recovered and is overtly recalled the first time it is sampled. Any attendant source information is also recovered, and a decision criterion is used to decide if there is enough source information to declare the item to be remembered. For the 6 *K* items, source information is sufficiently limited that it falls below the decision criterion, so a Know judgment is made. The 12 remaining traces contain more source information (represented by an "*R*" in Fig. 7). When they are sampled for the first time, the word is overtly recalled. In addition, because the amount of recovered source information falls above the decision criterion, a Remember judgment is made.

In all 4 of our experiments, Know judgments were less accurate than Remember judgments. However, Know judgments made with high confidence were as accurate as (or nearly as accurate as) Remember judgments made with high confidence. In Experiment 3, the mean RT data reflected these differences in strength, which is consistent with the random search model just outlined. However, the random search model makes more specific predictions than that. The model predicts that when cumulative recall is analyzed, both Remember and Know judgments should be characterized by a negatively accelerated exponential growth to asymptote, with the rate of approach to asymptote being slower for the weaker Know judgments compared to the stronger Remember judgments. The model further predicts that the rates of approach to asymptote should be similar once strength is equated by limiting the analysis to Remember and Know judgments made with high confidence. To test these predictions, the individual RTs that occurred during a recall period can be used to plot cumulative recall progress throughout the recall period. The cumulative plot shows the number of items recalled as a function of time spent recalling (which equals the sum of the RTs associated with the words that have been recalled thus far). When analyzed this way, the RTs are conceptualized as interresponse times (IRTs). Because the clock was stopped while the subject made confidence ratings and Remember/Know judgments, the recall time mainly reflects search time.

Fig. 8 shows the cumulative recall functions for Remember and Know judgments (summed across

participants) from the free recall condition of Experiment 3. Fig. 8A shows the functions for all Remember and Know judgments, and Fig. 8B shows the functions for high-confidence Remember and Know judgments. The curves drawn through the data show the least-squares fits of the standard 3-parameter exponential of the form $R = a * (1 - e^{-(t-c)/\tau})$, where R is the cumulative number of words recalled, a is the estimated asymptote, τ is mean recall latency (the parameter that governs rate of approach to asymptote), and c is an offset parameter that reflects the average time taken to type a word. Clearly, both Remember and Know judgments exhibit the typical exponential rise to asymptote that has long been known to characterize free recall (and that has been interpreted to reflect a search process from an episodic memory search set based on relative strength).

Table 5 shows the parameter estimates associated with the best-fitting exponential functions shown in Fig. 8. When all Remember and Know judgments are included in the analysis (R_{1-5} vs. K_{1-5}), Know judgments exhibit a somewhat slower rate of approach to asymptote (consistent with the fact that they reflect weaker memories). However, when the analysis is limited to high-confidence Remember and Know judgments – when the strength of Remember and Know judgments is equated (R_5 vs. K_5) – their rates of approach to asymptote are equated as well. These results are consistent with the basic RT analyses presented earlier. More to the point, this pattern of results is as it should be if both Remember and Know judgments reflect retrieval from the same episodic memory search set.

Although our theory holds that both Remember and Know judgments in free recall result from a consciously controlled search process and that they should exhibit similar recall dynamics when their strengths are equated, this should not be taken to mean that Remember and Know judgments are the same in all respects (even when they are equally strong). Remember and Know judgments do differ from each other in a theoretically significant way. However, according to this view, the difference between them is not that one reflects consciously controlled search and the other automatic memory. Instead, the difference is that some retrieved items are associated with source attributes (and receive Remember judgments), whereas others

Table 5
Parameter estimates in Experiment 3 for the free recall condition. The estimated asymptote is denoted by "a", "τ" is mean recall latency and "c" reflects the average time taken to type a word.

Parameter	R_{1-5}	K_{1-5}	R_5	K_5
a	1007.1	469.6	918.7	379.0
τ (s)	28.6	32.3	28.6	28.5
c (s)	2.7	3.5	2.4	3.0

are not (and receive Know judgments). This view seems quite similar to the position taken by Hamilton and Rajaram (2003), who wrote: "Thus, Remember and Know responses in free recall may be based on access to specific attributes vs. strength of item memory, respectively" (p. 66). It is also similar to the view espoused by Bodner and Lindsay (2003), who argued that, in conjunction with task demands, Remember and Know judgments in recognition memory are made based on the attributes of memory that are retrieved. If the source attributes that are retrieved help to solve the task at hand, a remember judgment is made; if not, a Know judgment is made (Gruppuso, Lindsay, & Kelley, 1997).

Recently, McCabe et al. (2010) presented evidence in favor of the idea that Know judgments in free recall reflect automatic memory. They found that dividing attention at study had no effect on the proportion of words from the study list that were recalled and given a Know judgment (0.14 in the full-attention condition; 0.13 in the divided-attention condition). By contrast, dividing attention at study selectively and dramatically reduced the proportion of words from the study list that were recalled and given a Remember judgment (0.35 in the full-attention condition; 0.16 in the divided-attention condition). They interpreted these results to mean that retrieval during free recall occurs either through a consciously controlled (memory search) process, yielding Remember judgments when successful, or through an automatic process, yielding Know judgments when items simply pop into mind without conscious effort. The same automatic process, when it occurs during recognition, theoretically gives rise to the feeling of familiarity.

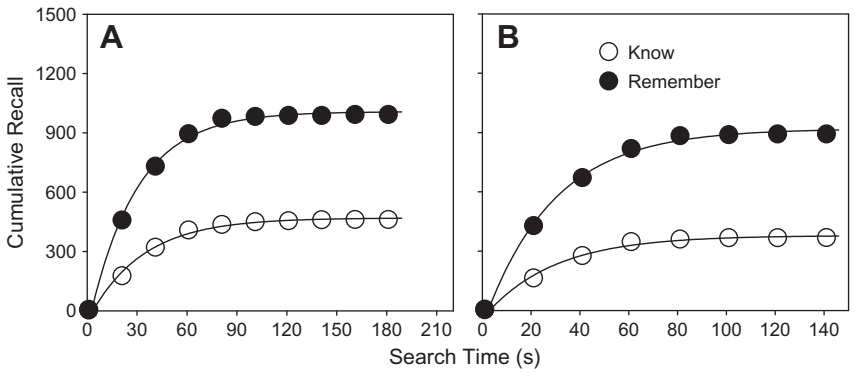

Fig. 8. Cumulative recall curves for the Free Recall condition of Experiment 3. Panel A shows the results for all Remember and Know judgments (i.e., collapsed across all levels of confidence), and Panel B shows the results for words recalled with high confidence only.

If Remember and Know judgments in free recall both reflect the outcome of a consciously controlled search process, then why did McCabe et al. (2010) find that dividing attention at encoding had no effect on the proportion of list items recalled with a Know judgment? One possibility is that dividing attention reduced the quality of all of the encoded traces such that some of the *K* items converted to unrecoverable ∼ items and some of the *R* traces converted both to *K* traces (encoded successfully enough to allow the item to be recovered, but without source information) and to ∼ traces. Because *K* traces are both lost and gained in this scenario, the overall effect might be no change in the proportion of list items receiving a Know judgment. The idea that traces initially supporting Remember judgments may later support Know judgments when memory has been weakened has been advanced in other contexts as well (e.g., Conway, Gardiner, Perfect, Anderson, & Cohen, 1997; Knowlton & Squire, 1995). Here, we add the further idea that, in free recall, traces that initially support a Know judgment may later convert into an unrecoverable traces.

One way to empirically differentiate between the consciously controlled vs. automatic accounts of Know judgments in free recall might be to divide attention at retrieval. If Know judgments reflect consciously controlled search, as we assume, then the rate of approach to asymptote should be slowed to the same extent as for Remember judgments. If they instead reflect automatic retrieval, then it seems reasonable to predict that Know judgments would be slowed to a much lesser extent than Remember judgments.

Remember/Know judgments in recognition

Our findings do not speak directly to Remember/Know judgments in recognition, but they do raise an intriguing possibility, one that might have direct relevance to a longstanding debate about the role of the hippocampus in recollection and familiarity. In studies of recognition memory, one common interpretation of Remember/Know judgments (based on dual-process theory) is that Remember judgments reflect recollection and Know judgments reflect familiarity. Another common interpretation (based on signal-detection theory) is that Remember judgments reflect strong memory and Know judgments reflect weaker memory. Although much evidence supports the latter interpretation (e.g., Wixted & Stretch, 2004), Wixted & Mickes (2010) found that Know judgments were associated with less source memory than Remember judgments even when equally strong memories were compared (a finding that our current research shows holds true of free recall as well). Thus, once memory strength is equated in terms of confidence and accuracy (unfortunately something that is rarely done), the evidence supports a distinction that goes beyond memory strength.

Wixted and Mickes (2010) assumed that high-confidence Know judgments mainly reflect familiarity-based decisions in recognition. Here, we consider the alternative possibility that while high-confidence Know judgments reflect the subjective experience of familiarity, that experience may be based on an underlying memory process that is more akin to recall than it is to processes that are ordinarily thought to underlie familiarity, such as perceptual fluency or automaticity. More specifically, our suggestion is that if high-confidence Know judgments in free recall reflect cue-dependent retrieval from episodic memory (like Remember judgments do), then there is no reason to assume that the same recall process does not occur for high-confidence Know judgments in recognition. This view accords with Tulving's (1985) view of cue-dependent recognition. For example, he wrote, "from the point of view of theory, and until such time as someone produces evidence to the contrary, we should assume that all retrieval is always cued" (p. 171). One need not fully embrace that strong view to accept the possibility that recognition sometimes consists of the cue-dependent retrieval of item-only information. Although the retrieval cue differs in the two tasks (in free recall, no explicit cue is provided; in recognition, a copy cue is provided), the end result may sometimes be the same, namely, the *recall* of item-only information.

Why, though, would item-only recall on a recognition memory task give rise to the subjective experience of familiarity – and to a Know judgment? One possible answer is that when a copy cue elicits the recall of item-only information in a recognition task, the subject may be aware that the recalled representation corresponds to the cue that retrieved it. The conscious awareness of this correspondence may be subjectively experienced as a strong sense of familiarity. Moreover, standard Remember/Know instructions would call for a Know judgment under these conditions because of the presence of item information and the absence of source information.

If this account is correct, then the subjective experience of recollection and familiarity would not align with the underlying memory processes giving rise to those experiences. If not, then the difference between two sides in a longstanding debate in the cognitive neuroscience literature may not be as great as it now seems. More specifically, a prominent view in the cognitive neuroscience literature is that the hippocampus selectively subserves recollection (e.g., Eichenbaum, Yonelinas, & Ranganath, 2007). One piece of evidence supporting this view is that activity in the hippocampus is reliably elevated for Remember judgments (compared to the activity for misses or correct rejections) but is not reliably elevated for Know judgments (e.g., Eldridge, Knowlton, Furmanski, Bookheimer, & Engel, 2000). Others (e.g., Squire, Wixted, & Clark, 2007) have interpreted this pattern to mean that, to detect activity in the hippocampus using fMRI, memory must be strong (indicated by high confidence and high accuracy). Because Know judgments typically reflect weak recognition memory (associated with low confidence and low accuracy, on average), elevated activity might not be detected for that reason alone.

Consistent with the latter interpretation, Smith, Wixted, and Squire (2011) recently found that when activity in the hippocampus was compared for high-confidence Remember and high-confidence Know judgments (both of which involved similarly high levels of old/new accuracy), elevated hippocampal activity was evident for both. This finding was interpreted to mean that the hippocampus supports strong familiarity as well as strong recollection.

However, if high-confidence Know judgments often reflect the outcome of an item-only *recall* process (even when memory is tested by recognition), it would mean that the hippocampus supports the subjective experience of familiarity, but it might nevertheless be true that the hippocampus supports a recall-like process (as opposed to a perceptual integration or perceptual fluency process). Moreover, this interpretation is consistent with the idea that, during encoding, the hippocampus serves an associative function, binding the studied item to contextual features (Eichenbaum et al., 2007). After all, it is the copy cue – together with contextual cues – that retrieve the item-only information in recognition memory. Sometimes, however, additional source information is not retrieved (at least not enough to warrant a Remember judgment), and that may be when (a) the subject experiences a strong sense of familiarity and (b) makes a Know judgment. Viewed in this light, the two sides in the debate over whether or not the hippocampus supports familiarity are not as far apart as is usually assumed.

We do not mean to suggest that the familiarity of an item on a recognition memory test always reflects the recall of item-only information. Our point instead is that Know judgments made with high-confidence and high accuracy may often be based on item-only recall, whether the test involves recall or recognition. Familiarity-based decisions that are generally made with lower confidence may be based primarily on other processes, such as perceptual fluency. Indeed, as others have suggested, there may be more than one kind of familiarity (Rugg & Curran, 2007; Rugg & Yonelinas, 2003), and this is what we are also suggesting here. In a review article critiquing the field of memory, Hintzman (2011) singled out the concept of familiarity as being too nebulous. About this, he wrote:

> "Familiarity is routinely invoked in formal and informal explanations of memory as though it were a concept with obvious meaning, but the term appears to mean more than one thing. This may be a case where the scientific adoption of a term from everyday life conveys explanatory power that is largely an illusion. The field could benefit from a careful analysis of the ways in which the concept of familiarity has been used." (p. 259)

Our findings using free recall raise the possibility that one form of familiarity may be more like recall than is ordinarily assumed.

Acknowledgments

This work was supported by Award Number R01MH082892 from the National Institute of Mental Health. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of Mental Health or the National Institutes of Health.

References

Bodner, G. E., & Lindsay, D. S. (2003). Remembering and knowing in context. *Journal of Memory and Language, 48*, 563–580.

Brainerd, C. J., & Reyna, V. F. (2010). Recollective and nonrecollective recall. *Journal of Memory and Language, 63*, 425–445.

Coltheart, M. (1981). MRC Psycholinguistic database. *Quarterly Journal of Experimental Psychology, 33A*, 497–505.

Conway, M. A., Gardiner, J. M., Perfect, T. J., Anderson, S. J., & Cohen, G. (1997). Changes in memory awareness during learning: The acquisition of knowledge by psychology undergraduates. *Journal of Experimental Psychology: General, 126*, 393–413.

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

Eichenbaum, H., Yonelinas, A. P., & Ranganath, C. (2007). The medial temporal lobe and recognition memory. *Annual Review of Neuroscience, 30*, 123–152.

Eldridge, L. L., Knowlton, B. J., Furmanski, C. S., Bookheimer, S. Y., & Engel, S. A. (2000). Remembering episodes: A selective role for the hippocampus during retrieval. *Nature Neuroscience, 3*, 1149–1152.

Gardiner, J., & Richardson-Klavehn, A. (2000). Remembering and knowing. In *The Oxford handbook of memory*. New York: Oxford University Press.

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

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*, 259–279.

Hamilton, M., & Rajaram, S. (2003). States of awareness across multiple memory tasks: Obtaining a "pure" measure of conscious recollection. *Acta Psychologica, 112*, 43–69.

Hintzman, D. L. (2011). Research strategy in the study of memory: Fads, fallacies, and the search for the "coordinates of truth". *Perspectives on Psychological Science, 6*, 253–271.

Ingram, K. M., Mickes, L., & Wixted, J. T. (2012). Recollection can be weak and familiarity can be strong. *Journal of Experimental Psychology-Learning, Memory, and Cognition, 38*, 325–339.

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

Jacoby, L. L., & Dallas, M. (1981). On the relationship between autobiographical and perceptual learning. *Journal of Experimental Psychology: General, 110*, 306–340.

Kintsch, W. (1970). Models for free recall and recognition. In D. A. Norman (Ed.), *Models of human memory*. New York: Academic Press.

Knowlton, B. J., & Squire, L. R. (1995). Remembering and knowing: Two different expressions of declarative memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 2l*, 699–710.

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

Mandler, G. (2008). Familiarity breeds attempts: A critical review of dual-process theories of recognition. *Perspectives on Psychological Science, 3*, 390–399.

McCabe, D. P., Roediger, H. L., & Karpicke, J. D. (2010). Automatic processing influences free recall: Converging evidence from the process dissociation procedure and remember-know judgments. *Memory & Cognition, 39*, 389–402.

McDermott, K. B. (2006). Paradoxical effects of testing: Repeated retrieval attempts enhance the likelihood of later accurate and false recall. *Memory & Cognition, 34*, 261–267.

Quamme, J. R., Yonelinas, A. P., Kroll, N. E. A., Sauve, M. J., & Knight, R. T. (2004). Recall and recognition in mild hypoxia: Using covariance structural modeling to test competing theories of explicit memory. *Neuropsychologia, 42*, 672–691.

Roediger, H. L., & Payne, D. G. (1985). Recall does not affect recall level or hypermnesia: A puzzle for generate/recognize theories. *Memory & Cognition, 13*, 1–7.

Roediger, H. L., & Thorpe, L. A. (1978). The role of recall time in producing hypermnesia. *Memory & Cognition, 6*, 296–305.

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

Rotello, C. M., & Zeng, M. (2008). Analysis of RT distributions in the remember–know paradigm. *Psychonomic Bulletin & Review, 15*, 825–832.

Rugg, M. D., & Curran, T. (2007). Event-related potentials and recognition memory. *Trends in Cognitive Sciences, 11*, 251–257.

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

Smith, C. N., Wixted, J. T., & Squire, L. R. (2011). The hippocampus supports both recollection and familiarity when memory is strong. *Journal of Neuroscience, 31*, 15693–15702.

Squire, L. R., Wixted, J. T., & Clark, R. E. (2007). Recognition memory and the medial temporal lobe: A new perspective. *Nature Review Neuroscience, 8*, 872–883.

Tulving, E. (1983). *Elements of episodic memory*. Oxford, England: Clarendon Press.

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

Tulving, E. (1972). Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), *Organization of memory* (pp. 381–403). New York: Academic Press.

Unsworth, N. (2007). Individual differences in working memory capacity and episodic retrieval: Examining the dynamics of delayed and continuous distractor free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*, 1020–1034.

Unsworth, N., Spillers, G. J., & Brewer, G. A. (2012). Evidence for noisy contextual search: Examining the dynamics of list-before-last recall. *Memory, 20*, 1–13.

Van Overschelde, J. P., Rawson, K. A., & Dunlosky, J. (2004). Category norms: An updated and expanded version of the Battig and Montague (1969) norms. *Journal of Memory and Language, 50*, 289–335.

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

Wixted, J. T., & Mickes, L. (2010). A continuous dual-process model of remember/know judgments. *Psychological Review, 117*, 1025–1054.

Wixted, J. T., & Squire, L. R. (2010). The role of the human hippocampus in familiarity-based and recollection-based recognition memory. *Behavioural Brain Research, 215*, 197–208.

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

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.