Available online at www.sciencedirect.com
SCIENCE DIRECT®
ACADEMIC PRESS
Journal of Memory and Language 49 (2003) 231–248
Journal of Memory and Language
www.elsevier.com/locate/jml

A dual-process account of the list-length and strength-based mirror effects in recognition[☆]

Melanie Cary* and Lynne M. Reder

Department of Psychology, Carnegie Mellon University, Pittsburgh, PA 15213, USA

Received 20 June 2002; revision received 30 September 2002

Abstract

Manipulating either list length (e.g., few vs. many study items) or encoding strength (e.g., one presentation vs. multiple presentations of each study item) produces a recognition mirror effect. A formal dual-process theory of recognition memory that accounts for the word-frequency mirror effect is extended to account for the list-length and strength-based mirror effects. According to this theory, the hit portions of these mirror effects result from differential ease of recollection-based recognition, and the false alarm portions result from differential reliance on familiarity-based recognition. This account yields predictions for participants' Remember and Know responses as a function of list length and encoding strength. Empirical data and model fits from four experiments support these predictions. The data also demonstrate a reliable list-length effect when several potential confounding factors are controlled, contributing to the debate regarding the effect of list length on recognition.

Keywords: Memory; Word recognition; Mirror effects; Encoding strength; List length

At one level of description, theories of recognition memory can be classified as either single- or dual-process. Some theorists posit that recognition performance is based on a single process, such as familiarity (e.g., Gillund & Shiffrin, 1984; Glanzer, Adams, Iverson, & Kim, 1993; Hintzman, 1988; McClelland & Chappell, 1998; Murdock, 1997; Shiffrin & Steyvers, 1997), while others posit that recognition performance is based on two processes, such as familiarity and recollection[1] (e.g., Jacoby, 1991; Jacoby & Dallas, 1981; Mandler, 1980; Reder et al., 2000; Yonelinas, 1994, 1999). Regardless of theoretical orientation, a complete theory of recognition memory must account for mirror effects (Glanzer & Adams, 1985). A mirror effect is said to occur when one experimental condition elicits more hits and fewer false alarms than another condition. The mirror effect that has probably received the most attention is the word-frequency mirror effect, in which words with a low normative frequency have a higher hit rate and a lower false alarm rate than words with a high normative

☆ This work was supported by Grants 5-T32-MH19983, 2-R01-MH52808, and T32-MH19102 from the National Institute of Mental Health. We thank Vincent Stretch and John Wixted for providing the Remember and Know data from their (1998) Experiment 1. We also thank Jason Arndt for his insightful comments on previous versions of this manuscript. Portions of these data were presented at the 42nd Annual Meeting of the Psychonomic Society, Orlando, FL.

* Corresponding author. Present address: Department of Psychology, University of Wisconsin-La Crosse, 335 Graff Main Hall, 1725 State Street, La Crosse, WI 54601, USA. Fax: +608-785-8433.

E-mail address: cary.mela@uwlax.edu (M. Cary).

[1] In contrast to current single- and dual-process models, Kelley and Wixted (2001) have recently proposed a model in which recollective (i.e., associative) information and familiarity-based (i.e., item) information are summed to produce one strength-of-evidence variable that is used to make a recognition decision.

doi:10.1016/S0749-596X(03)00061-5

frequency (e.g., Glanzer & Adams, 1985, 1990). Recently, Reder et al. (2000) proposed and found support for a formal dual-process account of the word-frequency mirror effect embedded within the Source of Activation Confusion (SAC) theory of memory. Using the same framework, they suggested accounts for the list-length and strength-based mirror effects, but did not test them. In this paper we examine whether such dual-process explanations of these two mirror effects are adequate.

The list-length effect refers to the finding that items from a longer list are recognized more poorly than items from a shorter list (e.g., Strong, 1912). Currently, there is debate as to whether this effect truly exists. Several studies have shown an effect of list length on recognition (e.g., Gillund & Shiffrin, 1984; Gronlund & Elam, 1994; Murnane & Shiffrin, 1991; Ohrt & Gronlund, 1999; Ratcliff, Clark, & Shiffrin, 1990; Ratcliff & Murdock, 1976; Strong, 1912), and the occurrence of a list-length effect has been widely accepted and considered a touchstone for models of recognition memory. However, some researchers have suggested that the list-length effect may be the result of confounds, such as retention interval, rather than list length (Dennis & Humphreys, 2001; Murdock & Kahana, 1993). To examine this alternative account of the list-length effect, Ohrt and Gronlund (1999) manipulated list length and controlled for retention interval, the number of items between study and test, the number of items scored, and the study position of tested items. Even with all of these controls, they found a reliable effect of list length on recognition, suggesting that the list-length effect is not simply an experimental artifact. In contrast, Dennis and Humphreys (2001) recently conducted a very controlled investigation of list length and found no reliable list-length effect. This issue will be addressed in our third experiment, in which we implement the same controls as Dennis and Humphreys and find a reliable effect of list length on recognition. Therefore, we consider the list-length effect to be a real recognition memory phenomenon and explore it as such.

The list-length effect is often examined in terms of overall performance measures, such as d' (e.g., Gillund & Shiffrin, 1984; Ohrt & Gronlund, 1999; Ratcliff et al., 1990). However when one examines the hit and false alarm patterns a *list-length mirror effect* is observed: There is a higher hit rate and lower false alarm rate for short lists than for long lists (see e.g., Murnane & Shiffrin, 1991; Ratcliff & Murdock, 1976).

Manipulating encoding strength also produces a mirror effect. Encoding strength is typically manipulated by varying either the presentation time for a single presentation of each item or the number of experimental presentations of each item. Overall, recognition is better for items from a strong encoding condition than for items from a weak encoding condition (e.g., Ratcliff et al., 1990). Underlying this effect is a *strength-based mirror effect* of more hits and fewer false alarms for the strong condition than for the weak condition (e.g., Murnane & Shiffrin, 1991; Stretch & Wixted, 1998).

The dual-process perspective has been shown to be beneficial in explaining the traditional word-frequency mirror effect (Joordens & Hockley, 2000; Reder et al., 2000). Additionally, this perspective has been used to predict conditions under which a word-frequency mirror effect should or should not occur. By manipulating factors that presumably affect the ease of recollection-based recognition, Joordens and Hockley (2000) created conditions under which a word-frequency mirror effect did not occur. Their results indicate that recollection, as well as familiarity, is important in producing the word-frequency mirror effect. Further, their results provide support for the influence of recollection in recognition.

A central goal of this paper is to test the adequacy of a dual-process account of the list-length and strength-based mirror effects in recognition. We do so by using a theory that has been developed to account for the word-frequency mirror effect. Although our goal is to make generalizations and conclusions concerning a class of dual-process theories, we focus on the SAC theory of memory (e.g., Reder et al., 2000) as a formally specified, computationally implemented example of this class. SAC accounts for patterns of Remember and Know responding as well as hit and false alarm patterns. The distinction between Remember and Know judgments refers to participants' classification of Old responses into those for which they can recollect a particular experience associated with the item (leading to a Remember response) and those for which the decision was based on a feeling of familiarity in the absence of recollection (leading to a Know response). From a dual-process perspective, it is especially useful to examine Remember and Know judgments in recognition memory tests, because they are presumably associated with the two processes that most dual-process theories claim people use to recognize a word, and, thus, they provide a converging measure or additional test of many dual-process theories.

In addition to offering a dual-process account of two mirror effects, we derive and test several predictions concerning the patterns of Remember and Know responses that underlie these effects. We test these predictions by presenting and simulating data from three new experiments and analyzing previously unreported data from a study by Stretch and Wixted (1998, Experiment 1). To our knowledge, this is the first study to examine Remember and Know responses as a function of list length or encoding strength. Moreover, we test for the presence of a list-length effect when several potential confounding factors are controlled.

A dual-process account of recognition memory

Like other dual-process memory theories, SAC incorporates the notion that recognition judgments can be based on either familiarity or recollection. Within SAC, recollection is conceived of as retrieving the trace of the encoded experience. When recollection-based recognition fails, the recognition judgment is based on familiarity. If the concept associated with the test item seems highly familiar, then a positive recognition may be made on that basis.

A formal, mathematical specification of SAC can be found in Reder et al. (2000); however, here we will outline the basic concepts. The model involves spreading activation within a network of nodes and links (i.e., associations) that vary in strength. The number of links and the strengths of the nodes and links vary as a function of experience. A node's strength translates into a resting level of activation. Nodes also exhibit transient changes in activation based on current stimulation from the environment or from spreading activation from associated nodes. The availability of a concept or an event depends on the current level of activation of the node that represents that item.

Within SAC, there are typically two consequences of encoding an item in memory. First, the strength of the *concept node* representing that item increases. Second, an *event node* representing the encoding event (e.g., where and when the word was encoded) is built. The event node is linked to the concept node and the node that represents the environmental context (e.g., the list context). The list *context node* has contextual features that are processed with the list items.

When a probe word is presented during a recognition test, both the corresponding concept node and the relevant context node are activated. The current activation of both nodes spreads to all associated nodes via the links that emanate from them. The amount of activation that spreads to any one associated node depends on the strength of the link to that node relative to the strength of all competing links from that source. Therefore, the amount of activation that spreads to an associated node from a node with many links will tend to be less than the amount spread to a node from one with only a few links.

Within this framework, there are two thresholds associated with recognition, an event threshold and a concept threshold. Recollection-based recognition responses occur when sufficient activation accrues at the associated event node for it to exceed the event threshold. In the absence of a recollection-based recognition, a familiarity-based recognition may occur when the concept node's activation is sufficient for it to exceed the concept threshold. This concept threshold is similar to the familiarity criterion posited by other dual-process perspectives (e.g., Yonelinas, 1994, 1999).

List-length mirror effect

Dual-process theories can naturally be extended to account for the list-length and strength-based mirror effects. The only necessary additional assumption is that the familiarity criterion that people adopt is affected by their perceptions of ease of recognition. Other researchers have advanced similar notions (e.g., Greene, 1996; Hirshman, 1995). Presumably, participants assume that some, typically half, of the words on the test list were on the study list, and, thus believe that they should produce a corresponding proportion of Old responses. Consequently, in test conditions in which recognition is perceived as more difficult, participants will use a lower familiarity criterion than in ones perceived to be easier.

From a dual-process perspective, recollection is an important factor in ease of recognition. When recollection is more difficult participants produce fewer recollection-based recognition responses, which can result in a relatively low proportion of Old responses. In this case, participants will tend to use a more liberal familiarity criterion. Thus, participants adjust their familiarity criterion in response to ease of recollection.

Within the framework of SAC, the list-length mirror effect is due to two factors that vary across lists of different lengths: the number of associations from the list context node and the concept threshold that participants adopt. When more items are studied on a list, there will be more associations from the list context node to the event nodes that represent the experience of studying each item on that list. At test, activation spreads from both the test item's concept node and the list context node. With more links emanating from the context node for a longer list, less activation will spread to any of its associated event nodes, thereby reducing the likelihood of an event node passing the threshold required for recollection. Consequently, there will be fewer recollection-based recognition responses for items from long lists than for items from short lists, producing the hit portion of the list-length mirror effect.

Because recollection-based recognition occurs less frequently after longer study lists, participants will rely more on familiarity to achieve more Old responses after studying long lists than after studying short lists. In doing so, participants will adopt a more liberal concept threshold, and thus make more familiarity-based false alarms for long than for short lists, producing the false alarm portion of the list-length mirror effect. We test two predictions derived from this account: Participants should make fewer recollection-based recognition hits and more familiarity-based false alarms as list length increases.

It is important to note that by expecting a list-length mirror effect in our data, we expect to find an effect of list length on overall recognition accuracy. Several memory models (e.g., Gillund & Shiffrin, 1984; McClelland & Chappell, 1998; Shiffrin & Steyvers, 1997) predict a list-length effect for recognition. In contrast, Dennis and Humphreys's (2001) recent context noise model of recognition memory predicts that list length does not affect recognition, and they found no reliable effect of list length. As noted earlier, they argue that previous list-length experiments allowed for confounding variables that produced spurious list-length effects.

Strength-based mirror effect

The strength-based mirror effect can also be explained by two factors that vary across the weak and strong encoding conditions: strength of memory traces and the familiarity criterion that participants adopt. Within SAC, each time an item is presented on the same study list, that item's concept node, study event node, and the link between them will get stronger. Thus, the activation levels and links for the concept and event nodes will be stronger for items studied in the strong condition than for items studied in the weak condition. At test, these relatively strong representations will make it easier for participants to recollect a study event for a target in the strong condition than one in the weak condition. As a result, there will be more recollection-based hits for the strong than for the weak condition, resulting in the hit portion of the strength-based mirror effect.

Parallel to the list-length mirror effect, participants will rely more on familiarity in the weak condition than in the strong condition. In other words, participants will adopt a more liberal concept threshold in the strong condition, producing the false alarm portion of the strength-based mirror effect. Thus, we test the predictions that participants should make fewer recollection-based recognition hits and more familiarity-based false alarms for the weak than for the strong condition.

In sum, the predictions for recollection-based recognition responses as a function of either list length or encoding strength come directly from pre-existing assumptions, and the predictions for familiarity-based recognition responses derive from the common assumption regarding a shift in the familiarity criterion (i.e., concept threshold) across conditions. As described in the next section, the Remember–Know paradigm provides a way to test the recollection and familiarity predictions for both mirror effects.

Remember–Know

The Remember–Know procedure (Gardiner, 1988; Tulving, 1985) requires participants to assess whether or not their Old response was based on a recollection of experiencing the item on the study list or whether it was based on a feeling of familiarity in the absence of recollection. Participants indicate the former by responding Remember and the latter by responding Know. This procedure has been used many times with considerable success in terms of separating these two types of responses (e.g., Gardiner, 1988; Gardiner & Java, 1990; Gardiner & Parkin, 1990; Rajaram, 1993). Moreover, different subjective experiences are associated with Remember and Know responses (Gardiner, Ramponi, & Richardson-Klavehn, 1998).

To assess the current predictions concerning recollection-based and familiarity-based responses, we examine Remember–Know data associated with the list-length and strength-based mirror effects. For both mirror effects, participants should make fewer Remember hits and more Know false alarms for the condition that yields poorer recognition (i.e, longer lists and weak encoding condition).

Experiment 1

Adopting a dual-process perspective allows us to make and test predictions about the patterns of Remember and Know responses that underlie the list-length mirror effect. We are unaware of existing Remember–Know data for this mirror effect. Participants in two experiments studied four lists of common nouns. Participants in a third experiment studied two lists of common nouns in an experiment that controlled for several potential confounding factors. At test, participants in all three experiments made Old–New and Remember–Know judgments. We expect to replicate the traditional list-length effect of poorer recognition for longer lists, and we predict a list-length mirror effect of fewer hits and more false alarms for longer lists. The dual-process theory we present also predicts fewer Remember hits and more Know false alarms for words from longer lists.

Method

Participants

The participants were 30 volunteer students from psychology courses at Carnegie Mellon University. In exchange for participating in one experimental session, they received credit towards one option of a course research requirement.

Materials and design

The word pool for targets and lures consisted of 320 one and two syllable nouns between four and eight letters in length with a Kucera and Francis (1967) frequency of at least six per million ($M = 42.85$). Presentation of stimuli and collection of responses was controlled by Macintosh computers using the PsyScope software (Cohen, MacWhinney, Flatt, & Provost, 1993).

List length varied within subjects across four study-test cycles. The four study lists consisted of 16, 32, 48, and 64 words. Each test list contained all of the studied words and an equal number of targets and lures. For example, a 48-item study list would be followed by a 96-item test list. For each participant, words were randomly assigned to be either a target or lure for one list-length condition. The words in each study and test list were presented in a random order.

Procedure

After reading and signing an informed consent form, participants completed four study-test cycles. The order of the different list-length conditions was randomized for each participant. Prior to each study list, participants were instructed to study each item in the list for a subsequent memory test. Each study item was presented for 1500 ms with an inter-stimulus interval of 500 ms.

Immediately after studying each list, participants performed a recognition test on it. For each test item, they made a yes–no recognition judgment. When participants indicated that an item had been on the study list, they also judged whether they "remembered" seeing the item on the study list or merely "knew" the item had been on the list. Participants received test instructions prior to each test list. These instructions informed them that they should make their yes–no judgments as quickly and accurately as possible. The Remember–Know portion of the test instructions closely followed those used by Knowlton and Squire (1995, p. 701). To establish that participants understood the Remember–Know task, they generated one example for each type of judgment prior to the first test list.

Results and discussion

Table 1 presents d', hits and false alarms as a function of list length. A one-way ANOVA indicated that recognition performance measured by d' was worse for words on longer lists, $F(3, 87) = 4.98$, $MSE = .42$, $p < .01$. This result replicates the general finding of a list-length effect for recognition.

Table 1
Recognition accuracy (d'), hit rates (Hits), and false alarm rates (FA) by list length for Experiment 1

Dependent measure	List length			
	16	32	48	64
d'	2.66	2.32	2.29	2.02
Hits	.78	.70	.70	.64
FA	.03	.06	.07	.08

The results also replicate the list-length mirror effect discussed in the Introduction. Participants made fewer hits and more false alarms for longer lists. Separate one-way ANOVAs for the hits and false alarms reveal that these effects of list length were reliable: $F(3,87) = 6.20$, $MSE = .02$, $p < .01$ and $F(3,87) = 6.03$, $MSE < .01$, $p < .01$ for hits and false alarms, respectively.

Fig. 1 displays the proportion of Remember and Know responses for hits and false alarms as a function of list length. We conducted separate two-way ANOVAs for Remember and Know responses with list length and whether the word had been studied as within-subjects factors. Overall, there were significantly more Remember responses for studied words ($M = .53$) than for unstudied words ($M = .01$), $F(1, 29) = 186.42$, $MSE = .09$, $p < .001$. Participants made fewer Remember responses to words from longer lists, $F(3, 87) = 6.17$, $MSE = .02$, $p < .01$. This main effect of list length was modulated by a reliable interaction of list length and whether the word had been studied, $F(3, 87) = 6.87$, $MSE = .02$, $p < .001$. The proportion of Remember responses for unstudied words did not differ across the four list-length conditions (all $M = .01$). In contrast, for words that had been studied, participants made fewer Remember responses for words on longer lists, one-way ANOVA $F(3, 87) = 6.59$, $MSE = .03$, $p < .001$. This result supports our prediction of fewer Remember hits for longer lists.

For Know responses, there were also significantly more responses for studied words ($M = .17$) than for unstudied words ($M = .05$), $F(1, 29) = 65.04$, $MSE = .01$, $p < .001$. There was a main effect of list length, $F(3, 87) = 5.23$, $MSE = .01$, $p < .01$, such that there were more Know responses for words from longer lists. There was not a reliable interaction of list length and whether the word had been studied, $F < 1$. As predicted, the proportion of Know false alarms increased with list length, mirroring the proportion of Remember hits.

SAC simulation

As an additional test of a dual-process account of the list-length mirror effect, we used the SAC model of memory to simulate how participants might perform in this experiment. The simulation generated the complete set of 16 data points (4 list lengths × 2 studied or unstudied × 2 Remember or Know) and was fit to the behavioral data aggregated over participants. Our simulation used the same functions and parameters as those used in the SAC simulation of the Remember–Know data for the word-frequency mirror effect (for details see Reder et al., 2000).

The current simulation produces a probability of responding Remember and a probability of responding

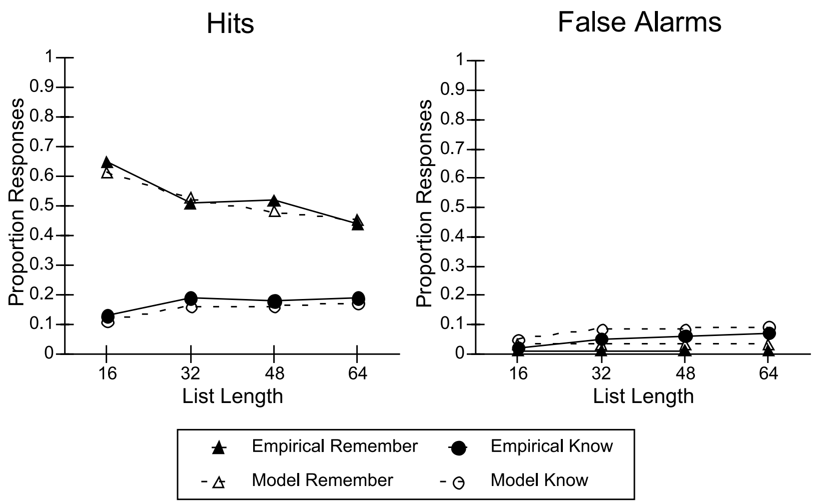

Fig. 1. Data and model fits for the proportion of Remember and Know responses for hits and false alarms as a function of list length, Experiment 1.

Know for each of the 8 experimental conditions. These probabilities are based on the activation values of the event nodes (for Remember responses) and concept nodes (for Know responses). The activation of an event node is influenced by its base-level strength and the amount of activation that spreads to it from the concept and list context nodes. The activation of a concept node is influenced by its base-level strength and input activation from reading the test word. The base-level strength of each node and the strength of each link increase with each experimental exposure and decrease over time. Pre-experimental experience with words is incorporated into the model by using normative word frequency to estimate pre-experimental base levels of activation and pre-experimental contextual associations for concept nodes.

The probability of responding Remember is essentially the probability that the event node's activation exceeds the event threshold. Because there is interdependence between Remember and Know responses, the estimated probability of responding Know is the probability of not responding Remember (i.e., one minus the probability of responding Remember) multiplied by the probability that the concept node's activation exceeds the concept threshold. Alternatively, $P(\text{Remember}) = r$, and $P(\text{Know}) = (1 - r) * k$, where r and k each represent the probability that the activation of the relevant node is sufficient to exceed the appropriate threshold. Thus, similar to other dual-process theorists (e.g., Jacoby, Jones, & Dolan, 1998; Yonelinas & Jacoby, 1995), we believe that the proportion of Know hits is not an accurate measure of familiarity.

To simulate the data, all parameter values except for the thresholds were the same as those used by Reder et al. (2000).[2] Based on our theoretical account of the list-length effect, there were four concept thresholds in the simulation: One for each list-length condition. We used a single event threshold for all conditions. Thus, the four concept thresholds and the one event threshold comprised the set of free parameters.

As shown in Fig. 1, our model provided an excellent fit to the data, yielding a Pearson r^2 of .99. The event threshold that was used for all list-length conditions was 74. The four concept thresholds for the four list lengths of 16, 32, 48, and 64 were 77, 72, 72, and 71, respectively. Thus, the concept thresholds were more liberal when the study lists were longer.

Experiment 2

Because the false alarm rate was quite low overall in Experiment 1, Experiment 2 was conducted to determine whether the same pattern of results would hold when performance was degraded. Thus, the goal of Experiment 2 was to replicate the results of Experiment 1 in a setting that would produce poorer recognition performance. This was done by adding a delay between each study and test list. The current dual-process framework predicts that a delay will lead to fewer recollection-based recognition responses due to decay of memory traces.

[2] In Reder et al. (2000) the value of the standard deviation for the concept threshold was different for Experiment 3 than for the first two experiments. Their Experiment 1 and 2 used a continuous recognition procedure, and their Experiment 3 used a more conventional paradigm. Therefore, this simulation used the standard deviation value from Experiment 3.

Moreover, participants will respond to this overall lower proportion of recollections by adopting a more liberal familiarity criterion, thus, making more Know responses than participants in Experiment 1. In other respects, the results of Experiment 2 should replicate those of Experiment 1.

Method

Participants

The participants were 18 volunteer students from psychology courses at Carnegie Mellon University. In exchange for participating in one experimental session, they received credit towards one option of a course research requirement.

Materials and procedure

All aspects of the experiment were the same as Experiment 1 except for the presence of an intervening task and the study presentation rate. Study items were presented for 1000 ms with an inter-stimulus interval of 500 ms. After each study list, participants worked on a word search puzzle for 5 min.

Results and discussion

Table 2 presents d', hits and false alarms as a function of list length. Due to the added delay between study and test and the reduced study time, we expected that participants in Experiment 2 would perform worse than those in Experiment 1. Visual inspection of Tables 1 and 2 indicate that, while the pattern of d', hits and false alarms is similar for the two experiments, participants in Experiment 2 performed more poorly across all four list-length conditions than those in Experiment 1.

A one-way ANOVA revealed that there was a reliable effect of list length on d', $F(3,51) = 3.25$, $MSE = .49$, $p < .05$, such that recognition performance was typically worse for words on longer lists. Generally, there was a list-length mirror pattern of fewer hits and more false alarms for longer lists. Separate one-way ANOVAs for hits and false alarms revealed that there was a reliable effect of list length on the proportion of false alarms, $F(3,51) = 3.76$, $MSE = .01$, $p < .05$. Although the proportion of hits decreased from the shortest to the longest list, the effect of list length on the proportion of hits was not statistically significant, $F(3,51) = 1.49$, $MSE = .02$, $p > .05$.

Table 2
Recognition accuracy (d'), hit rates (Hits), and false alarm rates (FA) by list length for Experiment 2

Dependent measure	List length			
	16	32	48	64
d'	1.95	1.51	1.59	1.23
Hits	.73	.66	.67	.64
FA	.13	.19	.18	.23

Fig. 2 displays the proportion of Remember and Know responses for hits and false alarms as a function of list length. We conducted separate two-way ANOVAs for Remember and Know responses with list length and whether the word had been studied as within-subjects factors. Overall, there were significantly more Remember responses for studied words ($M = .42$) than for unstudied words ($M = .04$), $F(1,17) = 53.84$, $MSE = .09$, $p < .001$. Participants made fewer Remember responses to words from longer lists, $F(3,51) = 3.08$, $MSE = .01$, $p < .05$. This main effect of list length was modulated by a reliable interaction of list length and whether the word had been studied, $F(3,51) = 3.70$, $MSE = .01$, $p < .05$. The proportion of Remember responses for unstudied words did not reliably differ across the four list-length conditions, one-way ANOVA $F(3,51) = 1.35$, $MSE < .01$, $p > .05$. In contrast, for words that had been studied, participants made fewer Remember responses for words on longer lists, one-way ANOVA $F(3,51) = 3.56$, $MSE = .02$, $p < .05$. These results replicate Experiment 1 and support the SAC prediction of fewer Remember hits for longer lists.

For Know responses, there were also significantly more responses for studied words ($M = .25$) than for unstudied words ($M = .14$), $F(1,17) = 12.36$, $MSE = .04$, $p < .01$. Replicating Experiment 1, there was a main effect of list length, $F(3,51) = 4.64$, $MSE = .01$, $p < .01$, such that there were more Know hits and false alarms for words from longer lists. There was no reliable interaction of list length and whether the word had been studied, $F < 1$. Notably, the pattern of Know false alarms mirrored the pattern of Remember hits.

Comparing the proportion of Remember and Know responses across Experiment 1 and 2, it appears that the differential delay between study and test affected performance as expected. Participants in Experiment 2 made fewer Remember hits and more Know false alarms than participants in Experiment 1. This pattern is consistent with the notion that participants in Experiment 2 relied more heavily on familiarity than participants in Experiment 1.

We also developed a SAC model of the Experiment 2 data. As with the simulation for Experiment 1, we modeled the complete set of 16 data points aggregated over participants, and the same five parameters were allowed to vary: one event threshold and four concept thresholds. Otherwise, the simulation used the same parameter values and functions as the Experiment 1 simulation.

As shown in Fig. 2, the SAC model produced a reasonable fit to the data, yielding a Pearson r^2 of .97. The event threshold that was used for all list-length conditions was 71. The four concept thresholds for the four list lengths of 16, 32, 48, and 64 were 67, 65, 64, and

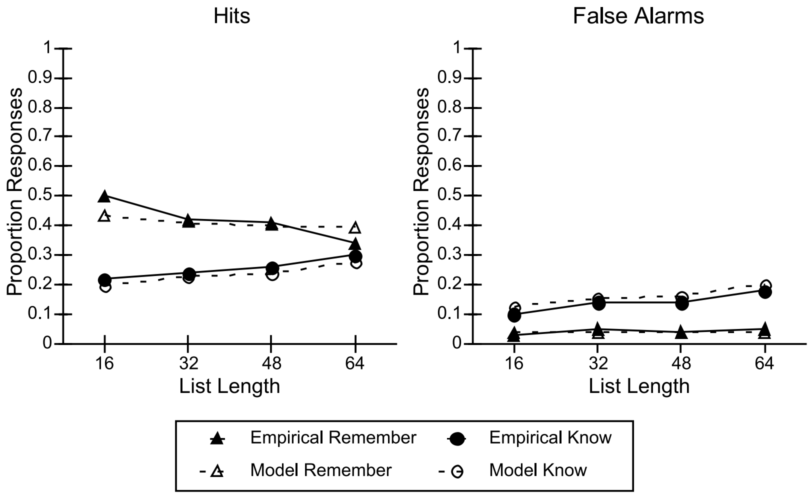

Fig. 2. Data and model fits for the proportion of Remember and Know responses for hits and false alarms as a function of list length, Experiment 2.

61, respectively. As with the simulation of the Experiment 1 data, the model's concept threshold is inversely related to list length.

Experiment 3

The results of Experiment 1 and 2 show an effect of list length on recognition memory and are consistent with a dual-process account. However, as mentioned in the Introduction, there is some debate in the literature as to whether an effect of list length really exists. For example, Dennis and Humphreys (2001) recently argued that when confounding factors are controlled there is no list-length effect in recognition. They suggested four possible confounds that could each lead to an artifactual list-length effect: retention interval, attention, rehearsal, and contextual reinstatement.

Retention interval is perhaps the most apparent factor. When the recognition test is presented either immediately after each study list or after an equal study-test delay the average retention interval is longer for the long list than for the short list, which could lead to poorer performance in the long list condition. This is also true when the long condition has more test items than the short condition. Dennis and Humphreys (2001) note that retention interval can be controlled by using either a retroactive or a proactive design. With a *retroactive design*, the delay between study and test is equated by (a) filling the time between study and test for the short list such that the total time from the first study item to the first test item is the same for each list length and (b) comparing recognition of the first items in the long list to recognition of words in the short list using test lists of the same length. With a *proactive design*, the delay is equated by (a) having an equal delay from the end of study to the beginning of test and (b) comparing only the last items in the long list to those in the short list using test lists of the same length.

A second potential confound is attention. Participants may give less attention to items at the end of a long study list than to items earlier on the list. If this occurs, items at the end of the long list will receive less attention than items on the short list. The influence of this type of effect would be exacerbated in the proactive design. Differential attention to study items can be limited by including an encoding task that requires participants to attend to each item. This will help ensure that participants process each study item.

Another possible confound that can occur with either the retroactive or proactive design is displaced rehearsals. When there is a study-test lag, participants may use part of the time during the lag to rehearse studied items. If rehearsal occurs, it could be more beneficial for performance in the short list condition. Typically, in both the retroactive and proactive designs, all of the studied items are tested in the short list condition and only a portion of the studied items is tested in the long list condition. Consequently, if participants rehearse after the short study list, any item that they rehearse will be tested, whereas if participants rehearse after the long study list, it is likely that they will rehearse both tested and untested items. Moreover, in the retroactive design, there is more time for participants to rehearse after the short study list than after the long study list. Dennis and Humphreys (2001) indicate that the potential influence of displaced rehearsals can be reduced by having an engaging filler task, by testing memory incidentally, and/

or by examining participants' performance for the targets that were presented earlier in each study list. By using this last control factor, the critical targets for the short list, as well as the long list, are a portion of the studied items. Consequently, if participants rehearse after either study list, they are likely to rehearse both examined and unexamined items.

The fourth potential confound presented by Dennis and Humphreys (2001) is the contextual reinstatement process. According to their *bind cue decide model of episodic memory* (BCDMEM), during a recognition test, participants cue memory with the test item and retrieve the set of contexts in which that item has been experienced. This retrieved context information is then compared to a context that the participant reinstates. When there is sufficient overlap in the retrieved and reinstated context information, participants will respond that the test item was on the study list. Consequently, according to BCDMEM, recognition performance will be accurate to the extent that participants reinstate an appropriate study context at test. Dennis and Humphreys assert that when the study-test lag is very short, participants may compare the retrieved context against either existing temporal cues or an end of list context, rather than reinstating an appropriate study context (e.g., a processing context). According to BCDMEM, an end of list context could provide a poor match to the retrieved contexts of targets presented early in the study list, such that participants who use an end of list context could miss more targets than participants who reinstate and use a more appropriate study context.

With regard to the list-length effect, a difference in the contextual reinstatement process for the short and long list conditions is most likely to occur when the retroactive design is used with a very short (e.g., 9 s) delay after the long list. The short delay for the long list condition might lead participants to use an end of list context during the recognition test, which could provide a relatively poor match to the context for the tested (early list) study items. Dennis and Humphreys (2001) further assert that in the short list condition, the intervening filler task should make any residual context information less useful, and, thus, participants will be more likely to reinstate an appropriate study context. Consequently, the (reinstated) context used at test could be more effective in the short list condition than in the long list condition, potentially resulting in better recognition of the short list items. This differential reinstatement of an appropriate study context can be minimized by having a filler task after both the short and long list and by having a distinctive encoding task (e.g., pleasantness rating) to implicitly encourage participants to attend to processing-based, rather than temporal, aspects of context.

In Experiment 3, we implemented controls for each of these four potential confounds by using both the retroactive and proactive control designs, equating the length of the test lists, maintaining a delay before each test list, including an encoding task, and examining performance for the targets that were presented earlier in each study list. Each participant completed four critical study-test cycles in a counterbalanced order: They studied and were tested after a short (20 item) list and a long (80 item) list under both the retroactive design and the proactive design. All test lists consisted of 40 items: 20 old and 20 new. During study, participants rated the pleasantness of each item, and at test they made Old–New and Remember–Know judgments. The retroactive condition was similar to Dennis and Humphreys's (2001) List-Length Experiment 2, with the notable exception that they used a short list of 40 items and we used a short list of 20 items. Thus, the difference between the short and long lists is greater here than in their experiment, where they found a null effect of list length on recognition.

Method

Participants

The participants were 40 volunteer students from psychology courses at Carnegie Mellon University. In exchange for participating in one experimental session, they received credit towards one option of a course research requirement.

Materials and design

The word pool was the same as in Experiments 1 and 2. List length and control design varied within subjects across four study-test cycles. Study list length was either short, consisting of 20 items, or long, consisting of 80 items. The control design was either retroactive or proactive. Table 3 displays the duration of the study list, the duration of the filler task, and which study items served

Table 3
Study list duration, filler task duration, and which study items served as targets by control design and list length for Experiment 3

	Proactive design		Retroactive design	
	Short	Long	Short	Long
Study list duration	60 s	240 s	60 s	240 s
Filler task duration	120 s	120 s	300 s	120 s
Target words	all 20	last 20	all 20	first 20

as targets for each of the four experimental conditions. In the proactive design, participants completed 2 min of a filler task between study and test, and the long condition test list contained the last 20 studied words along with 20 new words. In the retroactive design, participants completed 5 min of filler task after the short study list and 2 min of filler task after the long study list. The long condition test list contained the first 20 studied words along with 20 new words. Thus, the average study-test lag was equated for the short and long conditions. During the filler task participants solved a series of algebraic equations of various forms.

To avoid interrupting the critical timing in the experimental study-test cycles, participants also completed a sample, instructional, study-test cycle with a 20-item study list prior to the four experimental study-test cycles. Each of the five test lists contained 40 items: 20 targets and 20 lures. For the short list conditions and the sample all 20 study items were presented as targets. For each participant, words were randomly assigned to be either a target or a lure for one study-test cycle. The words in each study and test list were presented in a random order. At test, participants made both Old–New and Remember–Know judgments. As in Dennis and Humphreys (2001), the 20 target words in each condition can be viewed as being studied as two continuous blocks of 10 words each. To control for effects of differential rehearsal we analyze the hit data for the items that were presented in the first tested study block (i.e., the 10 targets that were presented earlier on the study list).

Procedure

After reading and signing an informed consent form, participants completed five study-test cycles: one sample cycle and four experimental cycles. The order of the four experimental conditions was counterbalanced across participants using a Latin square. The sample study and test lists were presented in a black font on a white background. The study and test lists for each experimental condition were presented in one of four colored fonts on a white background. Font colors were randomly assigned to experimental conditions for each participant. During study, participants rated the pleasantness of each item on a 4-point scale. Each study item was presented for 2500 ms with an inter-stimulus interval of 500 ms.

At the beginning of the sample cycle, participants received instructions on the rating task, which informed them that if they should happen to miss entering their rating while a word is displayed, then they should continue with the next word. After presentation of the sample study list, participants received instructions regarding the filler task and performed that task for approximately 2 min. Participants then received the same recognition task instructions as participants in Experiments 1 and 2 and performed the sample recognition test.

After the sample cycle, participants completed the four experimental cycles without interruption for instructions. Participants pressed the space bar to begin each study list. After each study list, the computer displayed the message "Perform the arithmetic task until you hear a tone" prompting participants to perform the filler task. At the end of the appropriate filler delay, a tone was played, and the computer displayed a ready message for 3 s prior to beginning the test list. Each participant participated in an individual session, rather than with a group, and an experimenter remained in the room throughout the entire session.

Results and discussion

Table 4 presents d', hits and false alarms as a function of list length and control design. Separate 2 (list length) $\times$ 2 (control design) ANOVAs were conducted for each dependent measure. For each of these measures, there was no reliable interaction of list length and control design, all $p > .05$. Importantly, there was a reliable effect of list length on d', $F(1,39) = 7.96$, $MSE = .25$, $p < .01$, such that recognition performance was worse for the long condition. This result provides support for the presence of a list-length effect, even when a variety of controls are incorporated into the experimental design.

There was also a list-length mirror pattern of fewer hits and more false alarms for the long condition than for the short condition. There was a reliable effect of list length on the proportion of false alarms, $F(1,39) = 6.78$, $MSE = .01$, $p < .05$. Although the proportion of hits was less for the long condition, this difference was not statistically significant, $F(1,39) = 2.65$, $MSE = .01$, $p > .05$.

Table 4
Recognition accuracy (d'), hit rates (Hits), and false alarm rates (FA) by list length and control design for Experiment 3

Control design and dependent measure	Short	Long	Mean
Retroactive design			
d'	2.44	2.23	2.34
Hits	.87	.83	.85
FA	.13	.15	.14
Proactive design			
d'	2.70	2.46	2.58
Hits	.87	.85	.86
FA	.07	.12	.10
Mean			
d'	2.57	2.34	2.46
Hits	.87	.84	.85
FA	.10	.14	.12

Recognition performance, measured by d', was worse with the retroactive design than with the proactive design, $F(1,39) = 7.46$, $MSE = .30$, $p < .01$. This effect is reflected in more false alarms for the retroactive design, $F(1,39) = 8.22$, $MSE = .01$, $p < .01$. There was no reliable effect of control design on the hit rate, $p > .05$.

Fig. 3 displays the proportion of Remember and Know responses for hits and false alarms as a function of list length and control design. We conducted separate three-way ANOVAs for Remember and Know responses with list length, control design, and whether the word had been studied as within-subjects factors. There were no reliable effects involving control design, ps $> .05$.

Overall, there were significantly more Remember responses for studied words ($M = .56$) than for unstudied words ($M = .04$), $F(1,39) = 193.76$, $MSE = .11$, $p < .001$. There was a reliable interaction of list length and whether the word had been studied, $F(1,39) = 6.61$, $MSE = .01$, $p < .05$. For words that had been studied, participants made fewer Remember responses for words in the long list condition ($M = .54$) than in the short list condition ($M = .58$), $t(39) = 2.01$, $p = .05$. The proportion of Remember responses for unstudied words did not reliably differ for the short and long list conditions (Ms $= .04$ and $.05$, respectively), $t(39) = 1.73$, $p > .05$. These results are consistent with the results of Experiments 1 and 2, where there were significantly fewer Remember hits for longer lists and the proportion of Remember false alarms did not vary as a function of list length. Thus, the results from these three experiments support our prediction of fewer Remember hits for longer lists.

There were also significantly more Know responses for studied words ($M = .29$) than for unstudied words ($M = .07$), $F(1,39) = 47.55$, $MSE = .08$, $p < .001$. Participants made slightly more Know responses in the long list condition ($M = .19$) than in the short list condition ($M = .17$). This difference was not statistically significant, and there was no reliable interaction of length and whether the word had been studied, ps $> .05$. The trend of more Know responses for the long list condition is consistent with the reliable effect of more Know

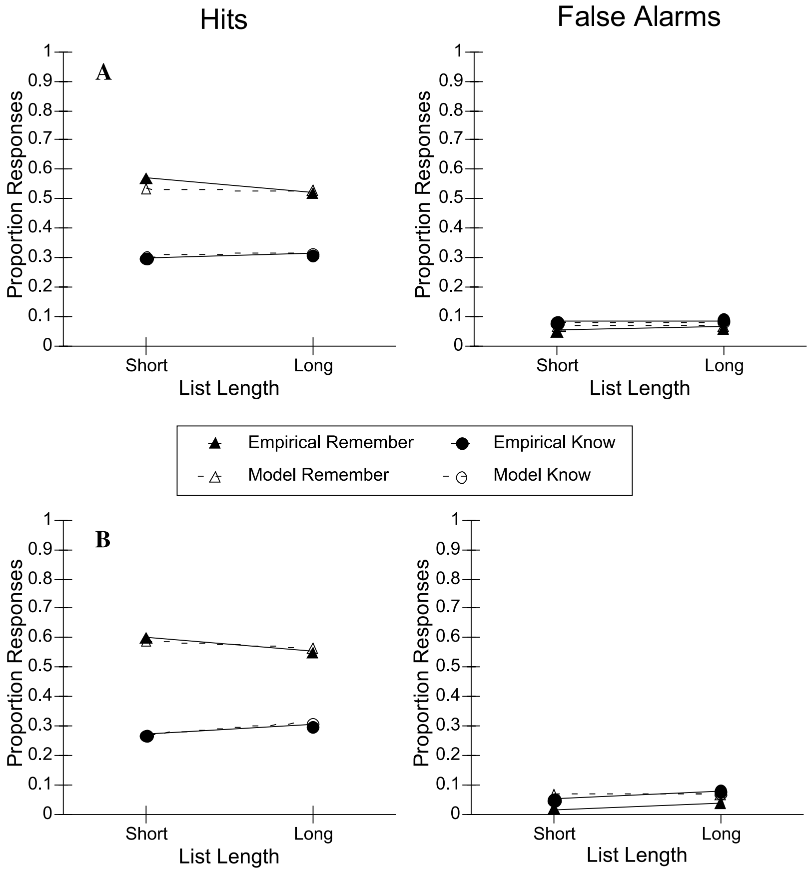

Fig. 3. Data and model fits for the proportion of Remember and Know responses for hits and false alarms as a function of list length and control design. (A) Retroactive design. (B) Proactive design.

responses for longer lists that was found in both Experiment 1 and 2.

We also developed a SAC model of the Experiment 3 data. As with the simulations for Experiments 1 and 2, we modeled the complete set of 16 data points (2 list length $\times$ 2 control design $\times$ 2 studied or not $\times$ 2 Remember or Know) aggregated over participants. There was a different concept threshold for each of the four test lists (i.e., retroactive short and long, proactive short and long). These four concept thresholds and one event threshold were allowed to vary. The concept threshold standard deviation from Experiments 1 and 2 of Reder et al. (2000) was used for all conditions. Otherwise, the simulation used the same parameter values and functions as the previous simulations reported here.

As shown in Fig. 3, the SAC model produced a good fit to the data, yielding a Pearson r^2 of .99. The event threshold that was used for all four conditions was 60. The concept thresholds for the short and long retroactive design conditions were 56 and 55, respectively. The concept thresholds for the short and long proactive design conditions were 57 and 56, respectively.

Summary

The data from three experiments support the Remember–Know predictions for the list-length mirror effect that were derived from the dual-process perspective of SAC. Participants produced fewer Remember hits and more Know false alarms for longer lists than for shorter lists. The results of Experiment 3 also indicate that a list-length effect can occur under very well-controlled experimental conditions. Next, we evaluate the Remember–Know predictions for the strength-based mirror effect.

Effects of encoding strength on Remember–Know responses

Stretch and Wixted (1998, Experiment 1) gathered the data necessary to test the dual-process predictions for the strength-based mirror effect: They manipulated encoding strength between lists and collected Remember–Know responses. In the weak condition each item was presented once, and in the strong condition each item was presented three times. Stretch and Wixted also varied normative word frequency within-lists. Because their intent was different than ours, they reported very little of the Remember–Know data. For methodological details see Stretch and Wixted (1998, Experiment 1).

Stretch and Wixted (1998) found both a word-frequency mirror effect and a strength-based mirror effect. We examined the proportion of Remember and Know responses made by their participants as a function of word frequency, encoding strength, and hits versus false alarms. Fig. 4 displays the proportion of Remember and Know responses for hits and false alarms as a function of encoding strength and word frequency. We conducted separate three-way analyses of variance (ANOVAs) for Remember and Know responses with word frequency, strength, and whether the word had been studied (hits vs. false alarms) as within-subjects factors. There were no reliable interactions involving both word frequency and strength for either Remember or Know responses, all Fs < 1.

Overall, there were significantly more Remember responses for studied words ($M = .54$) than for unstudied words ($M = .07$), $F(1, 35) = 376.35$, $MSE = .04$, $p < .001$. Participants made more Remember responses to words in the strong condition ($M = .34$) than to words in the weak condition ($M = .26$), $F(1, 35) = 21.16$, $MSE = .02$, $p < .001$. Importantly, this effect of strength was modulated by a reliable interaction of strength and whether the word had been studied, $F(1, 35) = 63.75$, $MSE = .01$, $p < .001$. As predicted, participants made substantially fewer Remember hits for words in the weak condition than for words in the strong condition, $t(35) = 6.49$, $p < .01$. In contrast, the proportion of Remember false alarms did not reliably differ for the strong and weak conditions, $t(35) = 1.25$, $p > .05$.

For Know responses, there were also significantly more responses for studied words ($M = .17$) than for unstudied words ($M = .11$), $F(1, 35) = 25.44$, $MSE = .01$, $p < .001$. There was a main effect of strength, $F(1, 35) = 12.07$, $MSE = .01$, $p < .01$, such that there were more Know hits and false alarms for words in the weak condition ($M = .17$) than for words in the strong condition ($M = .12$). There was no reliable interaction of strength and whether the word had been studied, $F(1, 35) = 1.90$, $MSE = .01$, $p > .05$. As predicted, participants made more Know false alarms in the weak than in the strong condition, $t(35) = 2.33$, $p < .05$.

Additional empirical test

Reder et al. (2000) predicted and found a Remember–Know mirror pattern for hits as a function of word frequency. We further analyzed Stretch and Wixted's (1998, Experiment 1) data to determine whether this mirror pattern was replicated in their data. Focusing first on the proportion of Remember responses, participants made significantly more Remember responses for low-frequency words ($M = .36$) than for high-frequency words ($M = .24$), $F(1, 35) = 108.65$, $MSE = .01$, $p < .001$. This effect was modulated by a reliable interaction of word frequency and whether the word had been studied, $F(1, 35) = 196.40$, $MSE = .01$, $p < .001$. Participants made substantially more Remember hits for low-frequency than for high-frequency words, $t(35) = 12.99$, $p < .001$, but they made slightly fewer Remember false alarms for low-frequency than for high-frequency words, $t(35) = 3.23, p < .01$. In contrast, there were significantly fewer Know responses to low-frequency words ($M = .11$) than to high-frequency words ($M = .18$), $F(1, 35) = 50.18$, $MSE = .01, p < .001$. This effect of word frequency on the proportion of Know responses was larger for hits than for false alarms, interaction $F(1, 35) = 5.61$, $MSE = .01, p < .05$. Thus, these results confirm the novel prediction from Reder et al. (2000) by replicating the Remember–Know mirror pattern: Participants made more Remember hits and fewer Know hits for low-frequency words than for high-frequency words.

SAC simulation

We modeled the 16 data points (2 low or high frequency × 2 weak or strong × 2 studied or unstudied × 2 Remember or Know) aggregated over participants. There were four free parameters: There was a concept threshold and event threshold for the weak condition and a concept threshold and event threshold for the strong condition. The use of two event thresholds, rather than one, is discussed in the General discussion. All other parameter values were the same as those used in the first two simulations reported here.

As shown in Fig. 4, the SAC model produced a good fit to the data, yielding a Pearson r^2 of .97. Consistent with our theoretical account of the strength-based mirror effect, the concept threshold for the weak condition ($T = 68$) was less than that for the strong condition ($T = 100$). This ordering was also true of the values derived for the event thresholds, which were 82 and 170 for the weak and strong conditions, respectively.

Summary

The data support SAC's qualitative Remember–Know predictions for the strength-based mirror effect: There were fewer Remember hits and more Know false alarms for words in the weak condition than for words in the strong condition. The results also replicate the Remember–Know word-frequency mirror pattern for hits that was identified by Reder et al. (2000). These findings support the utility of a dual-process perspective

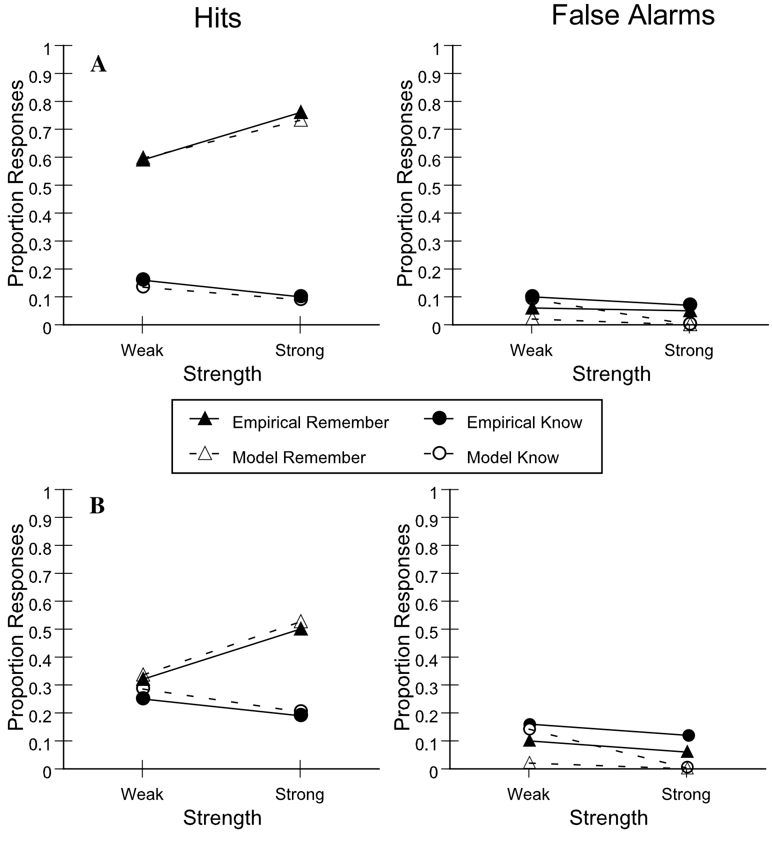

Fig. 4. Data and model fits for the proportion of Remember and Know responses for hits and false alarms as a function of strength and word frequency. (A) Low frequency. (B) High frequency.

for accounting for mirror effects in recognition and generating novel predictions.

General discussion

This study investigated the list-length and strength-based mirror effects in recognition from a dual-process perspective. Specifically, we used the SAC framework to provide dual-process accounts of these two mirror effects, tested predictions derived from these accounts with Remember–Know data, and modeled these data. We have also verified that the word-frequency Remember–Know effects predicted by SAC (Reder et al., 2000) were present in Stretch and Wixted's (1998, Experiment 1) data. To our knowledge, this is the first paper to investigate and simulate the influence of list length and encoding strength on Remember–Know judgments.

Additionally, we present empirical evidence that a list-length effect occurs even when several potential confounding factors are controlled. Murdock and Kahana (1993) and Dennis and Humphreys (2001) have asserted that the effects of list length that have been observed in the past have been due to confounding factors. Dennis and Humphreys controlled for each of four potential confounds (retention interval, attention, rehearsal, and contextual reinstatement) and found no statistically significant effect of list length on recognition. Experiment 3 of the present study controlled for these same factors, and found a significant list-length effect. It is likely that the list-length effect is real and that Dennis and Humphreys' manipulation was not strong enough to detect it.

Overall, the results are consistent with a dual-process theoretical account, realized via the SAC theory described earlier. Consistent with the assumption that the hit portion of the list-length mirror effect is due to less recollection-based recognition, participants in three experiments made fewer Remember hits, but more Know hits, for words from longer lists. Consistent with these data, Yonelinas and Jacoby's (1994) results from using the process-dissociation method indicate that the amount of recollection during recognition decreases as list length increases. Also, the assumption that the hit portion of the strength-based mirror effect is due to less recollection-based recognition in the weak condition than in the strong condition was supported: Participants in Stretch and Wixted's (1998) Experiment 1 made fewer Remember hits, but more Know hits, for weak than for strong targets. With regard to false alarms, we posit that the false alarm portions of the list-length and strength-based mirror effects are due to greater reliance on familiarity for longer lists and in the weak condition. Participants made more Know false alarms to lures in the weak condition and for longer lists. There was not a reliable difference in Remember false alarms as a function of either list length or encoding strength.

The SAC models produced good fits to the data. The few free parameters in each simulation were the concept and event thresholds. All other parameters were the values used for modeling the word-frequency mirror effect, and some of those parameter values had been retained from earlier models. Moreover, the ordinal patterns of the concept thresholds and the data were consistent with the assumption that people adopt different concept thresholds (i.e., familiarity criterions) as a function of the degree to which recollection contributes to recognition.

Comparisons with other accounts

With respect to the general mechanisms that underlie mirror effects, SAC shares some assumptions with other accounts. First, consider that within SAC the word-frequency mirror effect arises without a shift in the concept threshold, whereas a concept threshold shift underlies both the list-length and strength-based mirror effects. The notion that a criterion or threshold shift plays a role in some, but not all, mirror effects is not unique to the account presented here. Within the framework of a standard single-process signal detection perspective, Stretch and Wixted (1998) argue that there are two types of mirror effects that occur for distinct reasons. They suggest that one type of mirror effect occurs because participants shift their decision criterion across the different experimental conditions. They propose that a criterion shift plays a role in producing the strength-based mirror effect. Namely, participants are believed to adopt a more stringent decision criterion in the strong condition than in the weak condition. The other type of mirror effect suggested by Stretch and Wixted occurs when the criterion is fixed, but the lures in the different conditions differ in strength or familiarity. They argue that this is the case for the word-frequency mirror effect: Participants use the same decision criterion for low- and high-frequency words, but on average high-frequency lures are more familiar (i.e., stronger) than low-frequency lures.

The main purpose of Stretch and Wixted's (1998) study was to empirically examine whether a criterion shift underlies the word-frequency mirror effect. Their results provide evidence that participants do not shift their decision criterion as a function of word frequency. Their results also indicate that while people will make a criterion shift across test lists, they are reluctant to make a criterion shift on an item-by-item basis during a recognition test. From a dual-process perspective, we interpret these results, and the two types of mirror effects proposed by Stretch and Wixted, in terms of whether or not participants' shift their familiarity-based criterion. With this in mind, these results and the distinction between two types of mirror effects are consistent with the present theory.

Although Stretch and Wixted present a single-process signal-detection account of the two types of mirror effects, we favor a dual-process account. There is converging evidence in support of a dual-process theory of recognition memory (for a review see Yonelinas, 2002), including evidence from event-related potentials (e.g., Curran, 2000) and chemically induced temporary amnesia (Hirshman, Fisher, Henthorn, Arndt, & Passannante, 2002). Moreover, as discussed later, single-process signal-detection accounts are limited with regard to explaining Remember and Know responses.

Dual-process theories

The dual-process view presented here is similar to other dual-process perspectives. Notably, Joordens and Hockley (2000) also posit that the hit rate portion of the word-frequency mirror effect is due to less recollection-based recognition for high-frequency words because they have been seen in more contexts, and the false alarm portion is due to more familiarity-based responses for high-frequency words because they are more familiar. While Joordens and Hockley focused on the word-frequency mirror effect, in the General discussion they consider the strength-based mirror effect and sketch an account very similar to the one presented here. They propose that participants adopt a lower familiarity criterion for the weak condition than for the strong condition. Moreover, they conjecture that the familiarity criterion that a participant uses may be determined by the degree of recollection that occurs for that condition. As noted earlier, the current data and models are consistent with this conjecture.

The results of Jacoby et al.'s (1998) investigation of the effects of item repetition (i.e., strengthening) on recognition are congruent with the present dual-process account of the strength-based mirror effect. Using the Remember–Know paradigm and an exclusion paradigm, their results indicate that both recollection and familiarity increase with repetition. Moreover, they found that the increase in recollection for stronger items is reflected as an increase in the proportion of Remember hits; however, the increased familiarity associated with stronger items might not be reflected in participants' Know responses due to the nature of the relationship between recollection and familiarity.

The SAC framework is very similar to the dual-process views of Yonelinas (1994, 1999). Yonelinas posits a threshold recollection process and a signal-detection-based familiarity process. The basic assumptions of his model are the same as those of SAC; however, there are differences. With regard to recollection-based recognition, like Yonelinas, we believe that participants either recollect information about the encoding episode or they do not. However, in SAC the recollection-based recognition process is implemented based on a continuous distribution of event node activation values. When sufficient activation accrues at an event node to surpass the event threshold, then a person is able to recollect some information about that encoding episode. If the activation level does not exceed the threshold, then the person does not have a recollective experience.

A second distinction between SAC and Yonelinas' model is that they differ along the trace-aggregate dimension noted by Reder, Angstadt, Cary, Erickson, and Ayers (2002). Trace models attempt to explain behavior in terms of the principles governing the storage and retrieval of individual memory traces, whereas aggregate models operate at the aggregate level of a stimulus or manipulation class. Within this classification scheme, SAC is a trace model, and Yonelinas' model is an aggregate model. Yonelinas' model does not currently specify the underlying memory representations, including the nature of the representations that give rise to the posited threshold retrieval process.

Single-process theories

Several theorists have provided accounts of various mirror effects, including a set of single-process theories based around the assumption that people use likelihoods, rather than item strength, to discriminate between old and new stimuli (Glanzer et al., 1993; McClelland & Chappell, 1998; Shiffrin & Steyvers, 1997). Shiffrin and Steyvers's (1997) REM model and McClelland and Chappell's (1998) subjective-likelihood approach both attempt to account for the mirror effects considered here, as well as a variety of other memory phenomena. These two views are very similar to each other in many aspects and are quite different from SAC. Both REM and the subjective-likelihood model use noisy vector-based memory representations for items and are based on Bayesian likelihood ratios. To overcome limitations associated with using log-likelihood ratios, Sikström (2001) developed a variance theory implemented in a connectionist network that attempts to account for the same three mirror effects. Each of these theories provides an explanation for the list-length, strength-based and word-frequency mirror effects. However, in contrast to SAC, none of these models currently attempt to account for patterns of Remember and Know responses, and thus cannot currently provide as comprehensive a fit to the data as SAC.

Many other single-process theories have attempted to account for the list-length effect (i.e., poorer recognition for longer lists) and/or the finding that recognition performance is poorer after a weak than a strong encoding condition. For example, the ACT-R architecture has provided a formal retrieval-based single-process account of these two effects (Anderson, Bothell, Lebiere, & Matessa, 1998). Even though SAC is a dual-process theory and ACT-R is a single-process theory, they have similar strengthening and decay functions and both emphasize the role of associative interference. Within ACT-R, the poorer recognition performance for long lists and for weak lists is explained by two assumptions that are shared with SAC: There is more associative interference with longer lists than with shorter lists and stronger items have a higher base-level activation than weaker items. The ACT-R account of these effects was presented at the level of overall recognition performance (i.e., d') and currently does not account for patterns of Remember and Know responses.

A large class of single-process signal-detection models, known as global matching models, have attempted to account for the list-length effect (e.g., Gillund & Shiffrin, 1984; Hintzman, 1988; Murdock, 1982; Pike, 1984), but these accounts have been challenged by empirical findings. Some of these models (Gillund & Shiffrin, 1984; Hintzman, 1988; Pike, 1984) proposed to account for the list-length effect by positing that the variances of target and lure distributions increase as list length increases, such that longer lists result in noisier matching and therefore, poorer recognition than shorter lists. However, when Gronlund and Elam (1994) tested this increasing-variances account, it was not supported. An alternative account of the list-length effect (Murdock & Kahana, 1993) suggested that the effect is due to differential forgetting of items earlier in the list. Thus, there is more forgetting for longer lists due to the greater retention interval. However, a list-length effect occurs even when retention interval is controlled (e.g., Ohrt & Gronlund, 1999; current Experiment 3). In SAC, while delay affects recognition performance, the primary mechanisms for the effects of list length are the amount of activation spreading from the list context nodes and shifts in the concept threshold.

A benefit of SAC's dual-process perspective is that it provides good fits to and predicts aspects of the patterns of Remember and Know responses that underlie three types of mirror effects. It has been argued that variants of standard single-process signal-detection models can account for Remember–Know data (e.g., Donaldson, 1996; Hirshman & Master, 1997) by using two criteria within a signal-detection framework. In this view, people adopt an Old–New criterion along the familiarity continuum and a second Remember criterion farther out to the right on this continuum. This perspective suggests that there is no need to postulate a second process, such as recollection.

Although one could describe the patterns of data reported here within the two-criterion single-process perspective, there would be substantial limitations to these accounts. All such models that have been developed to account for patterns of Remember and Know data have two limitations (Gardiner, Richardson-Klavehn, & Ramponi, 1998). First, signal detection models of the Remember–Know paradigm have been post-hoc. These models have been developed to provide an account of extant data by placing response criteria in particular locations to account for the data. Thus, they do not explain why different factors influence the placement of these criteria nor do they predict patterns of Remember and Know responses. Second, they fail to account for subjective experiences associated with Remember and Know responses (Gardiner, Ramponi, & Richardson-Klavehn, 1998). These models have not explained how location or strength along a familiarity continuum can lead to qualitatively different experiences. (See the General discussion of Reder et al., 2000, for criticism specific to Donaldson's, 1996, argument in favor of a signal-detection account of Remember–Know responses.)

Challenge to the SAC model

Although the simulations provided good fits to the data, not every aspect of the simulations was as we would like. Recall that for the encoding strength simulation there were four free parameters: a concept threshold and an event threshold for each strength condition. There is no a priori reason, and it seems counterintuitive, to assume that participants would use different event thresholds for the weak and strong conditions. Yet, to obtain a good fit to the data, the simulation required two event thresholds. With only one event threshold, the simulation over-predicted the proportion of Remember hits and under-predicted the proportion of Know hits for the strong condition.

Why does the present encoding strength simulation require two event thresholds, rather than one? Including two event thresholds in the simulation compensates for a shortcoming in the computational model. Namely, that there is too much growth in the model with repeated presentations of an item. In the SAC simulations to date, each presentation of an item has the same effect on node and link strengths, regardless of whether the item has been seen recently. However, it is possible that participants exert less attentional or processing effort for a repeated presentation of an item that they have seen recently than they do for an item's first presentation. If this occurs, then repeated presentations of an item should produce less strengthening of nodes and links than the first presentation. In that case, the activation values for the strong condition should be lower than they are in the present model, potentially creating a circumstance in which one event threshold would be sufficient.

Conclusion

First, this study has provided additional evidence for the plausibility of a dual-process account of recognition memory. As a natural extension of an existing dual-process model of recognition called SAC, we derived predictions concerning phenomenological reports (i.e., Remember and Know responses) of participants as a function of list length and encoding strength. These predictions were supported in four experiments. By largely adopting assumptions and parameters that were incorporated in SAC to account for other memory phenomena, the model was able to provide quantitative fits to the data, supporting dual-process accounts of the list-length and strength-based mirror effects. Second, this study has provided additional evidence for the existence of a list-length effect for recognition memory by finding such an effect in a well-controlled experiment.

References

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

Cohen, J. D., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: A new graphic interactive environment for designing psychology experiments. *Behavioral Research Methods, Instruments, and Computers, 25*, 257–271.

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

Dennis, S., & Humphreys, M. S. (2001). A context noise model of episodic word recognition. *Psychological Review, 108*, 452–478.

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

Gardiner, J. M. (1988). Functional aspects of recollective experience. *Memory & Cognition, 16*, 309–313.

Gardiner, J. M., & Java, R. I. (1990). Recollective experience in word and nonword recognition. *Memory & Cognition, 18*, 23–30.

Gardiner, J. M., & Parkin, A. J. (1990). Attention and recollective experience in recognition memory. *Memory & Cognition, 18*, 579–583.

Gardiner, J. M., Ramponi, C., & Richardson-Klavehn, A. (1998). Experiences of remembering, knowing, and guessing. *Consciousness and Cognition, 7*, 1–26.

Gardiner, J. M., Richardson-Klavehn, A., & Ramponi, C. (1998). Limitations of the signal detection model of the Remember–Know paradigm: A reply to Hirshman. *Consciousness and Cognition, 7*, 285–288.

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

Glanzer, M., & Adams, J. K. (1985). The mirror effect in recognition memory. *Memory & Cognition, 13*, 8–20.

Glanzer, M., & Adams, J. K. (1990). The mirror effect in recognition memory: Data and theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*, 5–16.

Glanzer, M., Adams, J. K., Iverson, G. J., & Kim, K. (1993). The regularities of recognition memory. *Psychological Review, 100*, 546–567.

Gronlund, S. D., & Elam, L. E. (1994). List-length effect: Recognition accuracy and variance of underlying distributions. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 1355–1369.

Greene, R. L. (1996). Mirror effect in order and associative information: Role of response strategies. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 687–695.

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

Hirshman, E. (1995). Decision processes in recognition memory: Criterion shifts and the list-strength paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 302–313.

Hirshman, E., Fisher, J., Henthorn, T., Arndt, J., & Passannante, A. (2002). Midazolam amnesia and dual-process models of the word-frequency mirror effect. *Journal of Memory and Language, 47*, 499–516.

Hirshman, E., & Master, S. (1997). Modeling the conscious correlates of recognition memory. *Memory & Cognition, 25*, 345–351.

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

Jacoby, L. L., & Dallas, M. (1981). On the relationship between autobiographical memory and perceptual learning. *Journal of Experimental Psychology—General, 110*(3), 306–640.

Jacoby, L. L., Jones, T. C., & Dolan, P. O. (1998). Two effects of repetition: Support for a dual-process model of know judgments and exclusion errors. *Psychonomic Bulletin & Review, 5*, 705–709.

Joordens, S., & Hockley, W. E. (2000). Recollection and familiarity through the looking glass: When old does not mirror new. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 1534–1555.

Kelley, R., & Wixted, J. T. (2001). On the nature of associative information in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*, 701–722.

Kucera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University 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, 21*, 699–710.

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

McClelland, J. L., & Chappell, M. (1998). Familiarity breeds differentiation: A subjective-likelihood approach to the effects of experience in recognition memory. *Psychological Review, 105*, 734–760.

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

Murdock, B. B. (1997). Context and mediators in a theory of distributed associative memory (TODAM2). *Psychological Review, 104*, 839–862.

Murdock, B. B., & Kahana, M. J. (1993). List-strength and list-length effects: Reply to Shiffrin, Ratcliff, Murnane, and

Nobel (1993). *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*, 1450–1453.

Murnane, K., & Shiffrin, R. (1991). Interference and the representation of events in memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*, 855–874.

Ohrt, D. D., & Gronlund, S. D. (1999). List-length effect and continuous memory: Confounds and solutions. In C. Izawa (Ed.), *On human memory: Evolution, progress, and reflections on the 30th anniversary of the Atkinson-Shiffrin model* (pp. 105–125). Mahwah, NJ: Lawrence Erlbaum Associates.

Pike, R. (1984). A comparison of convolution and matrix distributed memory systems. *Psychological Review, 91*, 281–294.

Rajaram, S. (1993). Remembering and knowing: Two means of access to the personal past. *Memory & Cognition, 21*, 89–102.

Ratcliff, R., Clark, S. E., & Shiffrin, R. M. (1990). List-strength effect: I. Data and discussion. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*, 163–178.

Ratcliff, R., & Murdock, B. B., Jr. (1976). Retrieval processes in recognition memory. *Psychological Review, 83*, 190–214.

Reder, L. M., Angstadt, P., Cary, M., Erickson, M. A., & Ayers, M. S. (2002). A reexamination of stimulus-frequency effects in recognition: Two mirrors for low- and high-frequency pseudowords. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*, 138–152.

Reder, L. M., Nhouyvanisvong, A., Schunn, C. D., Ayers, M. S., Angstadt, P., & Hiraki, K. (2000). A mechanistic account of the mirror effect for word frequency: A computational model of remember–know judgments in a continuous recognition paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 294–320.

Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: REM–retrieving effectively from memory. *Psychonomic Bulletin & Review, 4*, 145–166.

Sikström, S. (2001). The variance theory of the mirror effect in recognition memory. *Psychonomic Bulletin & Review, 8*, 408–438.

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

Strong, E. K., Jr. (1912). The effect of length of series upon recognition memory. *Psychological Review, 19*, 447–462.

Tulving, E. (1985). Memory and consciousness. *Canadian Psychologist, 26*, 1–22.

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

Yonelinas, A. P. (1999). The contribution of recollection and familiarity to recognition and source-memory judgments: A formal dual-process model and an analysis of receiver operating characteristics. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*, 1415–1434.

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

Yonelinas, A. P., & Jacoby, L. L. (1994). Dissociations of processes in recognition memory: Effects of interference and of response speed. *Canadian Journal of Experimental Psychology, 48*, 516–534.

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