

Recollection Can Be Weak and Familiarity Can Be Strong

Katherine M. Ingram, Laura Mickes, and John T. Wixted
University of California, San Diego

The remember–know procedure is widely used to investigate recollection and familiarity in recognition memory, but almost all of the results obtained with that procedure can be readily accommodated by a unidimensional model based on signal-detection theory. The unidimensional model holds that remember judgments reflect strong memories (associated with high confidence, high accuracy, and fast reaction times), whereas know judgments reflect weaker memories (associated with lower confidence, lower accuracy, and slower reaction times). Although this is invariably true on average, a new 2-dimensional account (the continuous dual-process model) suggests that remember judgments made with low confidence should be associated with lower old–new accuracy but higher source accuracy than know judgments made with high confidence. We tested this prediction—and found evidence to support it—using a modified remember–know procedure in which participants were first asked to indicate a degree of recollection-based or familiarity-based confidence for each word presented on a recognition test and were then asked to recollect the color (red or blue) and screen location (top or bottom) associated with the word at study. For familiarity-based decisions, old–new accuracy increased with old–new confidence, but source accuracy did not (suggesting that stronger old–new memory was supported by higher degrees of familiarity). For recollection-based decisions, both old–new accuracy and source accuracy increased with old–new confidence (suggesting that stronger old–new memory was supported by higher degrees of recollection). These findings suggest that recollection and familiarity are continuous processes and that participants can indicate which process mainly contributed to their recognition decisions.

Keywords: recollection, familiarity, dual-process theory

A longstanding issue in the field of human memory is whether the ability to recognize an item as having been previously encountered is better characterized by a single-process model or by a dual-process model. Single-process models assume that recognition decisions are based on a continuous, unidimensional memory-strength variable that reflects a singular process like familiarity. Such models are often conceptualized in terms of the Gaussian unequal-variance signal-detection (UVSD) model (Egan, 1958). Dual-process models assume that recognition decisions are based on two different processes, recollection and familiarity (Atkinson & Juola, 1973; Jacoby, 1991; Mandler, 1980; Yonelinas, 1994). Recollection refers to the retrieval of contextual details associated with an item, and familiarity refers to a context-free sense of a prior occurrence generated by an item.

Within dual-process theory, a separate issue is whether recollection is better characterized as a dichotomous (high-threshold) process (Yonelinas, 1994) or as a continuous signal-detection

process (Wixted, 2007; Mickes, Wais, & Wixted, 2009). A dichotomous process either does or does not occur for a particular test item, whereas a continuous process occurs for every item to some degree. Most dual-process theories assume that familiarity is a continuous process, whereas recollection is a dichotomous process (e.g., Atkinson & Juola, 1973; Jacoby, 1991; Mandler, 1980). This idea is stated most clearly in the high-threshold/signal-detection (HTSD) model proposed by Yonelinas (1994). The HTSD model holds that recollection either fails or succeeds for a given test item. When recollection fails, recognition is based on a continuous familiarity process governed by an equal-variance signal-detection model. When recollection succeeds, an “old” decision is made with high confidence and high accuracy. According to an alternative dual-process account, recollection, like familiarity, is a continuous process that is associated with varying degrees of confidence and accuracy (Wixted, 2007). Although recollection and familiarity are both considered to be continuous variables, they still differ in that one process (recollection) involves the conscious retrieval of contextual information, whereas the other process (familiarity) does not. That is, the two processes differ in the content of the memory signal.

The three experiments reported here were designed to address both of these issues (i.e., single vs. dual-process accounts of recognition memory and dichotomous vs. continuous accounts of recollection). Each experiment used a variant of the remember–know procedure (Gardiner, Richardson-Klavehn, & Ramponi, 1997; Gardiner, 1988; Yonelinas, 2001). Typically, the remember–know procedure involves asking participants to provide remember

This article was published Online First October 3, 2011.

Katherine M. Ingram, Laura Mickes, and John T. Wixted, Department of Psychology, University of California, San Diego.

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. Some of this work was supported by National Institute of Mental Health Award R01MH082892.

Correspondence concerning this article should be addressed to John T. Wixted, Department of Psychology, 0109, University of California, San Diego, La Jolla, CA 92093-0109. E-mail: jwixted@ucsd.edu

judgments for old decisions that are based on recollection and to provide know judgments for old decisions that are based on familiarity. In the variant of this procedure that we used, participants were presented with a single scale that allowed them to express a degree of confidence for their remember or know judgment.

Although the remember–know procedure is widely used and continues to grow in popularity, the assumptions on which it is based have been increasingly challenged by studies suggesting that remember–know judgments are based on a single dimension of memory strength. The unidimensional interpretation is based on signal-detection theory (see Figure 1), and it holds that remember and know judgments reflect strong and weak memories, respectively (Donaldson, 1996). Memory strength in a unidimensional signal-detection model is often assumed to be based on a single process (such as familiarity), but it might also consist of the combination of continuous recollection and familiarity signals (Wixted & Stretch, 2004). Either way, the unidimensional account of remember–know judgments holds that these judgments reflect strong and weak memories, not recollection and familiarity.

Much evidence has accumulated in recent years suggesting that the unidimensional signal-detection account provides a more useful guide to thinking about remember–know judgments than does the standard dual-process account. For example, remember judgments are reliably associated with higher confidence, higher accuracy, and faster reaction times than know judgments, as a unidimensional strength account predicts (Dunn, 2004; Rotello & Zeng, 2008; Wixted & Stretch, 2004). In addition, clear criterion effects have been repeatedly observed for remember judgments (e.g., Rotello, Macmillan, Reeder, & Wong, 2005), and even behavioral dissociations that were once thought to be indicative of two processes are almost always easily reconciled with a unidimensional signal-detection interpretation (Dunn, 2004, 2008).

From a dual-process perspective, the challenge now is to offer an account of remember–know judgments that (a) is compatible with the large body of evidence that favors the unidimensional

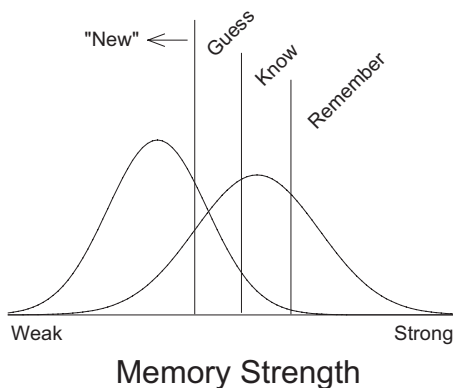


Figure 1. The signal-detection interpretation of remember–know–guess judgments. Recognition decisions are based on a continuous memory strength scale, and a remember judgment is made when the memory strength of a target or a lure exceeds a high criterion. A know judgment is made when memory strength only exceeds a lower criterion, and a guess judgment is made when memory strength only exceeds a still lower criterion.

signal-detection interpretation and (b) predicts findings that are not so easily reconciled with the unidimensional account. To that end, Wixted and Mickes (2010) proposed the continuous dual process (CDP) model illustrated in Figure 2, which, despite being a dual-process account, is purely signal-detection-based and naturally predicts many of the findings that have been advanced in support of the unidimensional signal-detection model. According to the CDP model, old–new decisions (including old–new confidence ratings) are based on a continuous, unidimensional memory strength signal consisting of the additive combination of continuous recollection and familiarity signals. Thus, its account of old–new recognition is mathematically identical to the longstanding UVSD model. This is an important consideration because the UVSD model has been differentially supported in numerous confidence-based receiver operating characteristic (ROC) experiments pitting it against the HTSD model (e.g., Heathcote, 2003; Jang, Wixted, & Huber, 2009; Rotello, Macmillan, Hicks, & Hautus, 2006; Starns & Ratcliff, 2008). However, the CDP model also assumes that when asked to do so, participants are able to query memory for the separable recollection and familiarity signals that comprise the aggregate memory strength variable. This is its key point of departure from the unidimensional, signal-detection-based interpretation of remember–know judgments advanced by Donaldson (1996).

According to the CDP model, after making an old–new decision based on aggregated memory strength, participants who are asked to make a remember–know judgment respond by selectively querying memory for evidence of recollection. Because the memory system is noisy and imperfect, a recollection signal of some magnitude is returned for every item (whether the item is a target or a lure). The recollection signal is assumed to be a continuously distributed Gaussian variable, and its average strength is higher for targets than for lures. That is to say, recollection is a signal-detection process (not a high-threshold process), which means that it requires a decision criterion. If enough recollection occurs for a test item (i.e., if the recollection signal exceeds a recollection-specific decision criterion), a remember judgment is made. If not, the participant then selectively queries memory for evidence of familiarity associated with the test item. If the familiarity signal exceeds the familiarity-specific decision criterion, a know judgment is made, and if not, the participant responds, “Guess” (if a guess option is included).

Like the unidimensional signal-detection model of remember–know judgments, the CDP model predicts that on average, remember judgments will be associated with higher confidence, higher accuracy, and faster old–new reaction times than know judgments (Wixted & Mickes, 2010). In addition, both models predict that remember judgments will exhibit criterion effects, such as the frequently observed positive correlation between remember hit and false alarm rates across participants (Wixted & Stretch, 2004). However, the CDP model predicts other findings that are less compatible with a unidimensional interpretation, a point that was illustrated in an experiment reported by Wixted and Mickes (2010). Participants in that experiment studied a list of targets, half in red or blue (color source) and half on the top or bottom of the screen (location source). On the subsequent recognition test, the targets were randomly intermixed with lures and, for each test item, participants rated old–new confidence using a 20-point rating scale (1 = 100% sure new, 20 = 100% sure old). For items rated

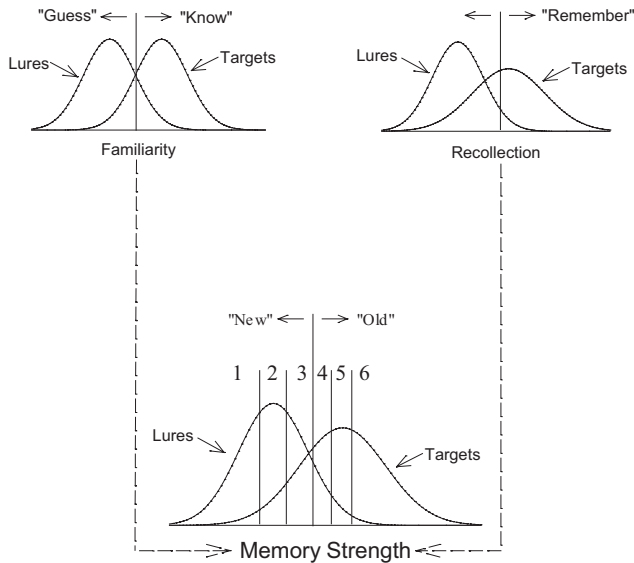


Figure 2. The continuous dual process (CDP) model of remember-know-guess judgments. To make an old-new decision, continuous recollection and familiarity signals are aggregated to form a continuous memory strength variable. The old-new confidence rating that is made depends on the highest criterion that is exceeded by the memory strength of the test item (the criteria for a 6-point confidence scale are shown). Thus, for an old-new decision, the CDP model is equivalent to the standard unequal-variance signal-detection model. To make a remember-know-guess judgment, the participant first queries memory for evidence of recollection and makes a remember judgment if the recollection signal exceeds the remember decision criterion. If not, the participant next queries memory for evidence of familiarity and makes a know judgment if the familiarity signal exceeds the know decision criterion (else, a guess judgment is made).

11 or higher (i.e., for words judged to be old), participants were first asked to make a remember, know, or guess judgment and were then asked to recollect criterial source details (“Was this word presented in red or blue? At the top or bottom of the screen?”).

At a coarse level of analysis, the results of this experiment were (as usual) compatible with the unidimensional signal-detection model. For example, confidence, old-new accuracy and source accuracy were all highest for Remember judgments, next highest for Know judgments and lowest for Guess judgments. This pattern of results is consistent with the idea that all three dependent variables provided measures of memory strength (Dunn, 2004; Wixted & Stretch, 2004). In addition, across participants, the remember hit rate was positively correlated with the remember false alarm rate, a common finding that is suggestive of a remember criterion (Wixted & Stretch, 2004). However, when the highest confidence ratings (20) for remember and know judgments were examined separately, a different story emerged. Old-new accuracy was approximately 91% in both cases, but source accuracy—an objective measure of recollection—was significantly greater for the remember-20 judgments (80% correct) than for the know-20 judgments (66% correct). Had source accuracy also been equated under these conditions, the results would have offered no reason to assume more than one dimension (or more than one memory process). In light of the substantial difference in source recollec-

tion for remember-20 and know-20 judgments, the results weigh in favor of the idea that remember-know judgments can be used to distinguish between decisions that are mainly based on recollection and those that are mainly based on familiarity. In particular, the know-20 judgments appear to be based primarily on familiarity, thereby accounting for the low level of source accuracy despite the high level of old-new accuracy.

Although the results favored a two-dimensional (dual-process) interpretation of remember-know judgments, a unidimensional model (one that need not assume two processes) could account for these findings if old-new accuracy for high-confidence remember judgments was, in truth, greater than old-new accuracy for high-confidence know judgments, but the difference was not detected because performance was approaching the ceiling. If old-new accuracy were greater for remember-20 judgments, then source accuracy should be (and was) greater as well. This interpretation is bolstered by the fact that in other studies reported by Wixted and Mickes (2010), old-new accuracy for high-confidence remember judgments was higher than old-new accuracy for high-confidence know judgments, even though the difference was not always significant.

A more compelling pattern of results would involve a comparison between remember judgments made with relatively low confidence and low old-new accuracy versus know judgments made with higher confidence and higher old-new accuracy. A unidimensional account predicts that if old-new accuracy is higher for high-confidence know judgments, compared with low-confidence remember judgments, then source accuracy should be higher for the know judgments as well. By contrast, the two-dimensional CDP model makes the opposite prediction. That is, the CDP model predicts that source accuracy for remember judgments made with lower confidence and lower old-new accuracy will be higher than source accuracy for know judgments made with higher confidence and higher old-new accuracy. Results like these would not only be inconsistent with a unidimensional account but would also be inconsistent with a threshold recollection account because remember judgments should not be made with varying degrees of confidence and accuracy (only high confidence and high accuracy). According to the CDP model, remember judgments made with varying degrees of old-new confidence should be associated with varying degrees of source recollection accuracy (because recollection is a continuous process).

Experiment 1

In most studies that collect both confidence ratings and remember-know judgments for each old decision, remember judgments are mainly associated with high confidence (e.g., 6 on a 6-point scale), and know judgments are associated with lower confidence (e.g., 5 or less on a 6-point scale). This pattern accords with the unidimensional signal-detection view (according to which remember judgments reflect strong memory, and know judgments reflect weak memory), and it also accords with the idea that recollection is a threshold process (one that yields high-confidence remember judgments whenever the threshold is exceeded). However, from the point of view of the CDP model, one reason why this pattern might be commonly observed is that participants are often admonished not to make Remember judgments unless they definitely remember source details. Sometimes, participants are warned that if

they make a remember judgment, their memory for source details will be tested. This warning is designed to ensure that they are not making a remember judgment when they cannot recollect contextual details (e.g., Yonelinas, 2001). Interpreted in terms of the CDP model, instructions like these may have the effect of encouraging participants to set a high remember criterion. In that case, only items associated with *strong* recollection would receive remember judgments, and such items would also be strong in terms of old–new memory strength (and would also receive a high-confidence old decision for that reason). Lesser degrees of recollection (i.e., below-criterion recollection) would be associated with know judgments, and this would be true whether the know judgments were made with low or high confidence.

If the CDP model is correct, it should be possible for participants to provide valid remember judgments when recollection is of lesser strength (e.g., when participants are reasonably sure, but not completely sure, that they are recollecting source details). In that case, remember judgments might be associated with lower levels of old–new confidence, even though they would be indicative of some level of source recollection. In agreement with this idea, previous studies have shown that remember judgments can be associated with lower levels of confidence when instructions are used to manipulate the remember criterion (e.g., Rotello et al., 2005). However, source recollection was not tested in these experiments, so it could be argued that low-confidence remember judgments were not valid (e.g., Parks & Yonelinas, 2007). The CDP model predicts that for any level of old–new confidence, remember judgments should be associated with higher levels of source accuracy than are know judgments. Moreover, high-confidence know judgments should be associated with higher old–new accuracy, but lower source recollection accuracy, than remember judgments made with lower levels of confidence. This pattern is not anticipated either by a single-process account or by a dual-process account that assumes threshold recollection.

Participants

There were 23 undergraduate participants (16 women, 7 men) from the University of California, San Diego (UCSD) who received course credit for participating in the experiment.

Materials and Design

Participants were asked to study a list of words for a later memory test. The word pool from which the list items were drawn consisted of 682 three-to-seven letter words generated from the medical research council Psycholinguistic Database (Coltheart, 1981), with the additional constraints of ratings between 500 and 700 for familiarity, imaginability, and concreteness. For each participant, 272 words were randomly selected from the word pool, with 128 words assigned as targets and 128 words assigned as lures. The remaining 16 words were randomly divided between targets and lures and used during the practice phase. The experiment was run using an E-prime program (Version 2.0; www.pschnet.com; Psychology Software Tools) to display the instructions and stimuli to participants.

Procedure

Participants signed a consent form, and then they were read instructions before they completed a short practice session to train

them on the procedure. Next, they proceeded to the study phase. During this phase, 128 target words were presented for 2.5 s at a time on the screen, in either red or blue and on the top or the bottom of the screen. The word list was balanced for both of these source attributes. In the test phase, the 128 targets were randomly intermixed with 128 lure words that were not on the study list. The words were tested one at a time, and each word was presented on the screen along with a 20-point rating scale. A response of 1–10 was used to indicate their confidence that the word was new, while a response of 11–20 was used to indicate their confidence that the word was old (where 1 indicates the highest confidence that the word was new and 20 indicates with highest confidence that the word was old). This scale, displayed in Figure 3, differs from previous rating scales in that old ratings with a confidence of 16–20 could be made on a familiarity scale or a recollection scale (i.e., 16F, 17F, 18F, 19F, and 20F, along with 16R, 17R, 18R, 19R, and 20R), where F means familiar and R means remember. As has been done in some prior studies (e.g., Montaldi, Spencer, Roberts, & Mayes, 2006), we exchanged *know* with *familiar* in an attempt to reduce confusion between the colloquial and experimental use of *know* (see McCabe & Geraci, 2009). This scale was used to provide a visual indication that there was no requirement for remember judgments to be made with high confidence only. Participants made old–new decisions by clicking on the 20-point rating scale using a mouse.

Participants were told that responses of 20R and 20F should be equal in strength. That is, they should both indicate 100% certainty that the word was on the list; the only difference is the ability to recall details about its prior occurrence. A lower rating of familiarity or recollection was meant to indicate decreased confidence that the word was on the list and decreased familiarity or recollection. For example, a rating of 16R would be used to indicate fairly high confidence that the word was previously presented, with the additional thought that perhaps something of its prior occurrence could be recalled. A rating of 16F, on the other hand,

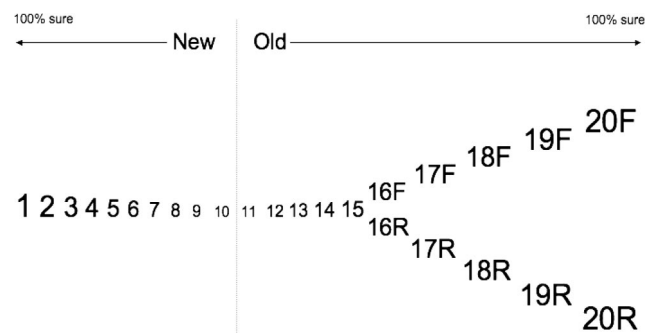


Figure 3. The 20-point rating scale used in Experiments 1 and 3. A rating of 1 corresponds to the highest level of confidence that the word is new (not on the study list). A rating of 20 corresponds to the highest level of confidence that the presented word was on the study list (old) and is further broken down by a simultaneous remember (R) versus familiar (F) judgment. Ratings of 16–19 correspond to varying levels of confidence and also require an R–F judgment. With regard to the continuous dual process model shown in Figure 2, a rating of 20R would be made if aggregated memory strength exceeded the highest confidence criterion on the memory strength axis and if the recollection signal exceeded the remember criterion on the recollection axis.

would indicate the same level of confidence that the word was previously presented, but without any recollection of contextual details associated with its prior occurrence. Ratings of 11–15 were not subdivided between recollection and familiarity because pilot testing indicated that source recollection for these ratings would not be sufficiently above chance to distinguish between them. For any word rated 11 or higher (i.e., for any word declared to be old), participants made a source decision by indicating whether the word was previously presented in red or blue and on the top or the bottom of the screen.

Results

Overall, 34% of the targets and 6% of the lures received a confidence rating of 16–20 (either recollection or familiarity). For remember judgments, the hit and false alarm rates were .14 and .01, respectively (93% correct), and the average confidence rating was 19.2. For familiar judgments, the hit and false alarm rates were .20 and .05, respectively (80% correct), and the average confidence rating was 18.5. Thus, in the aggregate, the results exhibit the typical strength pattern in that remember judgments were made with higher confidence and higher accuracy than were familiar judgments. In addition, the remember hit rate was positively (and significantly) correlated with the remember false alarm rate, $r(20) = .61$, which suggests differences in the placement of the remember criterion across participants. Findings like these, which have often been observed before, are compatible with the single-process signal-detection-based interpretation of remember–familiar judgments, but they are also compatible with the CDP model.

Participants made many high-confidence remember (20R) and familiar (20F) judgments, but the remainder of their ratings was distributed across the scale such that no one rating in the range of 16 through 19 received a large number of responses. Thus, we collapsed the data across 16F–19F (henceforth denoted F_{16-19}) and 16R–19R (denoted R_{16-19}) and compared familiarity-based and recollection-based decisions made with moderate confidence (i.e., F_{16-19} and R_{16-19}) to familiarity-based and recollection-based decisions made with confidence ratings of 20 (denoted F_{20} and R_{20}). Across all participants and summed across targets and lures, 433 observations fell into the F_{16-19} category, and 256 observations fell into the F_{20} category. In addition, 166 observations fell into the R_{16-19} category, and 270 observations fell into the R_{20} category. Eleven of the 23 participants made at least one response in all four of the relevant categories (F_{16-19} , F_{20} , R_{16-19} , and R_{20}), whereas 12 of the 23 did not. Thus, we first analyze the data from the 11 who provided a complete set of data (culminating in a state-trace analysis, which requires responses in all four categories) and then consider the remaining 12 participants separately.

The main question of interest was whether source accuracy for lower confidence remember responses would be lower or higher than the source accuracy for higher confidence familiar responses (even if old–new accuracy exhibited the opposite pattern). Source accuracy was first computed separately for each attribute (e.g., correct color judgments divided by correct plus incorrect color judgments). Because the accuracy scores for the two source attributes did not differ significantly and did exhibit similar trends, they were averaged together to reduce error variance. Figure 4A

shows that old–new accuracy was much higher for F_{20} responses (95%) than for R_{16-19} responses (81%). If both old–new accuracy and source accuracy were a function of a singular memory strength variable, then the large difference in old–new accuracy (from R_{16-19} to F_{20}) should be associated with higher source accuracy for F_{20} responses as well. Instead, although the difference was not significant, the trend was in the opposite direction. That is, source accuracy was higher for R_{16-19} responses (67%) than for F_{20} responses (58%). The interaction between response category (R_{16-19} vs. F_{20}) and type of memory test (old–new vs. source) was significant, $F(1, 10) = 6.75$, $p = .027$. Considered individually, source accuracy for R_{16-19} responses was significantly greater than chance, $F(1, 10) = 8.75$, but source accuracy for F_{20} responses was not, $F(1, 10) = 1.12$. This pattern of results corresponds to the pattern predicted by the CDP model.

We next compared accuracy scores for the two levels of remember judgments (i.e., R_{16-19} vs. R_{20}) for these 11 participants. That is, we compared performance across two levels of subjective recollection. Figure 4B shows that both old–new accuracy and source accuracy were lower for R_{16-19} (81% and 67%, respectively) than for R_{20} (97% and 82%, respectively). A paired t test comparing the old–new accuracy of R_{16-19} and R_{20} was significant, $t(10) = 3.01$, $p = .013$. Similarly, a paired t test comparing the source accuracy of R_{16-19} and R_{20} was significant, $t(10) = 2.76$, $p = .020$. This result is consistent with the idea that recollection is a continuous process. That is, source memory was significantly above chance for both R_{16-19} responses and R_{20} responses, but it was significantly greater for R_{20} than for R_{16-19} responses.

Figure 4C shows a similar analysis performed on F_{16-19} versus F_{20} judgments. This analysis compares performance across two levels of subjective familiarity. Old–new accuracy for F_{20} responses was significantly greater for F_{16-19} responses, $t(10) = 4.67$, $p = .001$, which is consistent with the widely accepted idea that familiarity is a continuous process. However, source accuracy was not significantly different between F_{16-19} (55%) and F_{20} (58%). Although source accuracy was above chance, as is typically the case for know judgments, it was not significantly greater than chance in this case. Averaged across the F_{16-19} and F_{20} categories, source accuracy was greater than .50 for seven participants, was equal to .50 for one participant, and fell below .50 for three participants.

A significant interaction of the kind shown in Figure 4A would offer compelling evidence against a unidimensional account if the proportion correct scores for old–new accuracy and source accuracy provided interval scales of measurement. To investigate whether a unidimensional model could accommodate these data when only an ordinal measurement scale is assumed, we next performed a state-trace analysis of the data from these 11 participants (Bamber, 1979; Dunn, 2008). In this context, a state-trace analysis involves plotting one dependent measure thought to be sensitive to variations in recollection against another dependent measure thought to be sensitive to variations in familiarity. The different levels of each dependent measure might be obtained by manipulating two independent variables that theoretically affect the two processes in different ways. If both dependent measures actually reflect a singular unidimensional memory strength variable, then the state-trace plot will vary monotonically across levels of the independent variables (i.e., as one dependent measure in-

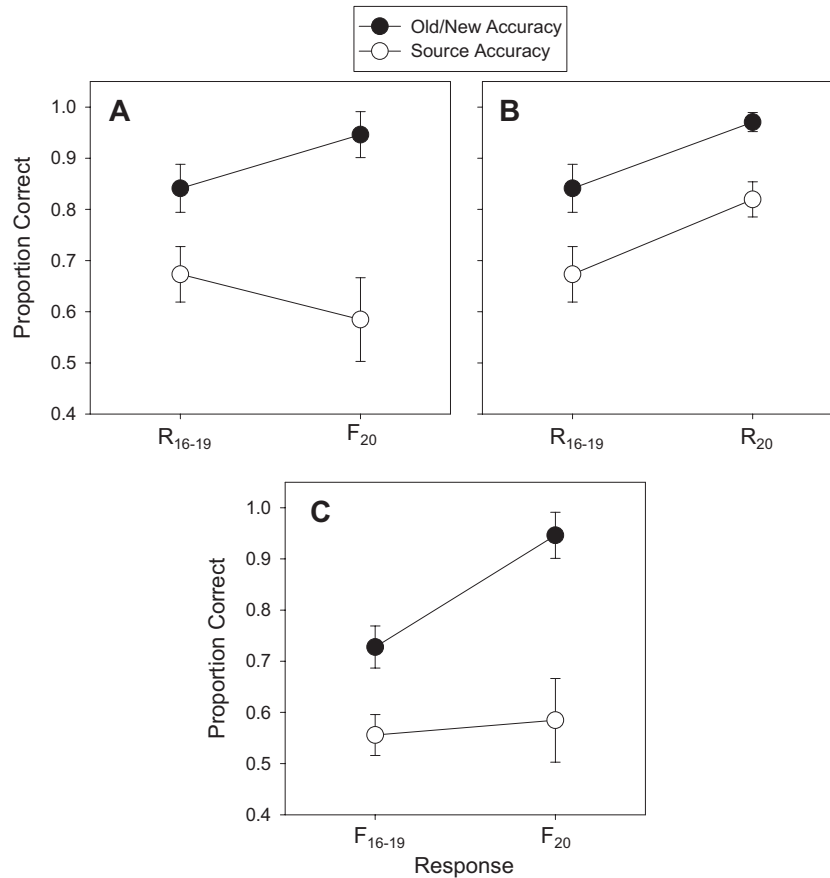


Figure 4. Proportion correct for old–new and source decisions in Experiment 1 for R_{16–19} and F₂₀ responses (A), R_{16–19} and R₂₀ responses (B), and F_{16–19} and F₂₀ responses (C). R = remember; F = familiar.

creases, the other will increase as well). If the measures are differentially sensitive to the effect of manipulating the independent variables, then the state-trace plot will be characterized by a nonmonotonic function (Bamber, 1979; Newell & Dunn, 2008). Somewhat counterintuitively, the existence of an interaction like that shown in Figure 4A does not guarantee a nonmonotonic state-trace plot.

The state-trace analysis relevant to our study involves plotting source recollection accuracy (a measure sensitive to recollection) against old–new accuracy (a measure sensitive to familiarity as well as recollection) for remember and familiar judgments made with different levels of confidence. Thus, two different kinds of judgments were used instead of manipulating two different independent variables. If the points on the state-trace plot follow a single, monotonically increasing function, the results would be consistent with the idea that remember and know judgments are based on a single-process. The state-trace plot for these 11 participants is displayed in Figure 5. The points on this plot do not appear to exhibit a monotonically increasing trend but instead appear to reflect separate increasing trends (one for remember judgments and the other for familiarity judgments). These data could reflect nonsystematic deviations from a monotonically increasing function, in which case they would not be inconsistent with a unidimensional model. However, the deviations are in the

direction predicted by the CDP model, and the consistency of that pattern is investigated in the next two experiments.

The results discussed above pertain to the 11 participants who provided responses in all four of the relevant response categories (and for whom a state-trace plot could be plotted), but what about the remaining 12 participants? Of these 12, five did not provide any ratings in the R_{16–19} category, but they did provide responses in the F_{16–19}, F₂₀, and R₂₀ categories. This pattern is predicted by a threshold recollection model because, according to that model, recollection is not associated with degrees of confidence (whereas familiarity is). An alternative interpretation is that these participants were only willing to say remember when they were certain of recollection (i.e., they had a high remember criterion). If the remember criterion is high, then a remember judgment will be given only when recollection is strong, and strong recollection will generally yield the highest level of old–new confidence. A high remember criterion would be reflected in a low remember hit and a low remember false alarm rate. In agreement with this idea, the average remember hit rate for these five participants was only 5% (far lower than the 20% remember hit rate of the 11 participants considered above), and their average remember false alarm rate was essentially 0% (lower than the 2% remember false alarm rate of the 11 participants considered above) because only one remember false alarm was made between the five participants. Thus,

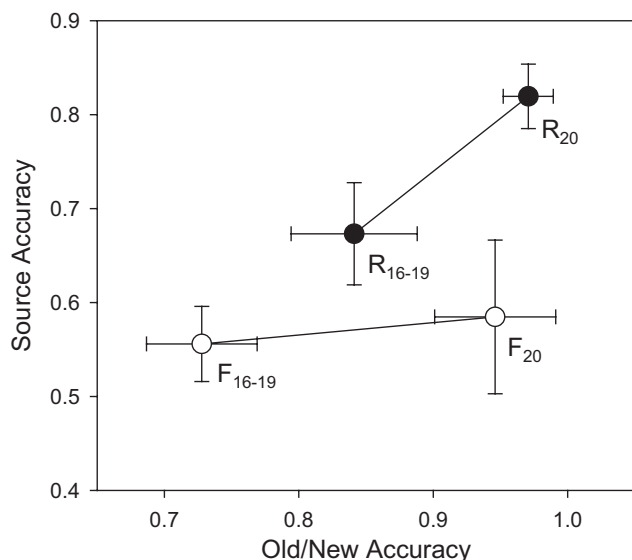


Figure 5. State-trace plot for Experiment 1 showing the mean old–new and source accuracies for F_{16-19} , F_{20} , R_{16-19} , and R_{20} . R = remember; F = familiar.

these participants appear to have been very conservative about declaring recollection.

According to the CDP model, a reluctance to declare recollection does not necessarily mean that recollection occurred only when R_{20} judgments were given. Instead, it means that recollection had to be strong before a remember judgment was made and that lesser degrees of recollection were therefore associated with familiarity judgments. The average source accuracy for the R_{20} responses made by these participants was 84% correct, whereas their average source accuracy for F_{20} responses was 64% correct. The F_{20} source accuracy scores were greater than 50% correct for all five of these participants. Thus, by the binomial, their source accuracy for F_{20} responses was significantly greater than chance. This result is consistent with the idea—inherent in the CDP model—that recollection is a continuous process and that familiar judgments indicate that not enough recollection occurred to make a remember judgment (Wais, Mickes, & Wixted, 2008). It might be argued that the above-chance source memory associated with F_{20} responses (both in the 11 participants considered earlier and in the five participants currently under consideration) reflects “unitized familiarity,” but it seems fair to say that this result is not predicted by process-pure models and that such models would have been differentially supported had source accuracy for F_{20} judgments fallen at chance levels (instead of at significantly above-chance levels).

Although the performance of these five participants accords with the CDP model, their performance can also be reconciled with a unidimensional memory strength model. That is, for the three categories of performance for which they supplied data (F_{16-19} , F_{20} , and R_{20}), source accuracy (52%, 64%, and 84%, respectively) tracked old–new accuracy (80%, 94%, and 98%, respectively). Thus, one need not assume a two-dimensional dual-process model to account for their performance, even though it is also true that their performance can be understood in terms of the CDP model.

Two other participants also clearly used the scale as a strength scale. One supplied confidence ratings in the range of 16–20 almost exclusively for familiarity judgments (with the exception of a single 18R response), and the other supplied confidence ratings in the range of 16–20 only for recollection judgments. Such performance is obviously consistent with a single-process signal-detection model.

Finally, the five remaining participants did not yield a pattern that would help distinguish between any of the models under consideration. Three were very conservative in their use of the upper end of the confidence scale. These three participants provided very few responses (10 or less) distributed across ratings of 16–20, making it difficult to analyze their remember and familiar judgments in a theoretically meaningful way. Finally, the two remaining participants had missing cells of data that seemed to reflect scale biases.

Overall, the results suggest that many participants use remember and familiar judgments in a way that accords with the CDP model, but some participants may simply use a rating scale as a memory strength scale (in accordance with the single-process signal-detection account).

Experiment 2

The results of Experiment 1 suggest that recollection and familiarity may be distinct, continuous dimensions of memory (including the recollection dimension). Experiment 2 was conceptually identical to Experiment 1, but we used a more traditional 6-point confidence scale and tested a larger number of participants.

Participants

There were 56 UCSD undergraduates (40 women, 16 men) who received course credit in exchange for their participation.

Materials and Design

The only difference from Experiment 1 was the old–new confidence rating scale. The rating scale used in Experiment 2 had six levels of confidence, with a response of 1–3 indicating different levels of confidence that the word was new, and a rating of 4–6 indicating different levels of confidence that the word was old (where 1 indicates the highest confidence that the word was new, and 6 indicates the highest confidence the word was old). Old ratings of 4–6 were further parsed by familiarity and remember options (4F, 5F, and 6F; 4R, 5R, and 6R). The rating scale is illustrated in Figure 6.

Procedure

The procedure was identical to that in Experiment 1 in every way, except that participants used a 6-point rating scale. As such, in the instructions it was emphasized to participants that responses of 6F and 6R should be equal with regard to memory strength. The only difference between familiarity and recollection was the ability to recall details about the given word’s prior occurrence. A lower rating of familiarity or recollection indicated decreased confidence that the word was on the list and decreased familiarity or recollection. For example, a rating of 5R would indicate that the word was probably on the study list and that some source details could

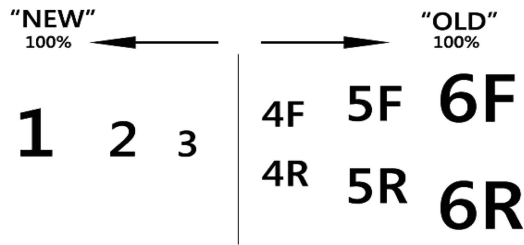


Figure 6. Rating scale used in Experiment 2. A rating of 6 corresponds to the highest level of confidence that the word was on the study list (old) and is further broken down by a remember–familiar judgment. Ratings of 4 and 5 reflect lower levels of confidence and also require a remember–familiar judgment. A rating of 1 reflects the highest level of confidence that the word was not previously studied (new). R = remember; F = familiar.

be recollected. A rating of 5F would also mean that the word was probably on the list, but no source details could be recollected. As before, participants were encouraged to use a recollection option if they recollected criterial source details (i.e., the color and/or location of the word at study) or any other source details (e.g., thoughts or images associated with the word at study). The participants were free to use the scale as they saw fit (i.e., they were not asked to use all of the scale options). For words that were declared to be old, participants next made two source (color and location) decisions.

Results

Collapsed across confidence ratings of 4, 5, and 6, the remember hit and false alarm rates were .24 and .04, respectively (85% correct), and the familiar hit and false alarm rates were .33 and .16, respectively (67% correct). Thus, overall accuracy was lower than it was in Experiment 1, but the typical strength pattern was still evident. It was also evident in the confidence ratings. The average confidence rating for remember judgments was 5.5, whereas the average confidence rating for familiar judgments was 4.8. The correlation between the remember hit rate and the remember false alarm rate across participants was significant, $r(53) = .59$. Thus, once again (and as usual), the aggregate data are consistent with a single-process strength account.

For consistency with the previous experiment, ratings of 4–6 for familiar and remember responses are denoted F_4 , F_5 , F_6 , R_4 , R_5 , and R_6 , respectively. Ratings of 4 were made infrequently, and source accuracy for ratings of F_4 and R_4 were both at chance (.52 and .51, respectively). Thus, our analyses focus on ratings of R_5 , F_5 , R_6 , and F_6 . Summed across participants, the number of observations falling into these four categories was 424, 1,040, 1,273, and 979, respectively. As in Experiment 1, we first analyzed data from participants who had at least one rating response in each of these four categories (because a state-trace analysis could be performed on such data). Forty participants satisfied these criteria. The data from the remaining 16 participants are considered separately.

Once again, the comparison of most interest involved relatively weak recollection (R_5) versus strong familiarity (F_6). As in Experiment 1, the accuracy scores for color and location did not differ significantly and exhibited similar trends, so they were averaged together to reduce error variance. Figure 7A shows that the old–new accuracy of the 40 participants who produced the full set of

data was higher for F_6 responses than for R_5 responses, 88% versus 83%. However, source accuracy for R_5 responses (67%) was greater than for F_6 responses (58%). Although neither individual test yielded a significant difference, the interaction between type of response (R_5 vs. F_6) and type of test (old–new vs. source) was again significant, $F(1, 39) = 5.45$, $p = .025$. This interaction was predicted by the CDP model but is not predicted either by threshold-recollection-based dual-process models or by single-process models.

The next analysis focused on recollection-based decisions (R_5 and R_6) for these 40 participants and is shown in Figure 7B. The difference in old–new accuracy between R_5 and R_6 responses (83% vs. 95%) was significant, $t(39) = 4.43$, $p < .001$, as was the difference in source accuracy for R_5 and R_6 responses (65% and 78%), $t(39) = 4.21$, $p < .001$. For both R_5 and R_6 responses, source accuracy was significantly above chance, $t(39) = 4.81$ and 16.3, respectively. From a dual-process perspective, these results again suggest that recollection occurs in degrees. That is, the participants subjectively indicated experiencing recollection to different degrees, and their objective source recollection performance is consistent with their subjective reports.

The same analyses were conducted on the familiarity responses for these 40 participants. Figure 7C shows old–new and source accuracies for familiarity responses for confidence ratings of 5 and 6. The old–new accuracy for F_6 responses was significantly greater than for the F_5 responses, $t(39) = 5.62$, $p < .001$. Source accuracy for F_5 and F_6 responses (58% and 60%, respectively) did not differ significantly, but overall source accuracy for F_5 and F_6 was greater than chance, $t(39) = 3.13$ and 3.92, respectively. These results are again consistent with the idea that recollection is a continuous process and that know (or, in this case, familiar) judgments indicate that not enough recollection occurred to make a remember judgment.

Figure 8 shows the state-trace plot of the old–new and source accuracy data for remember and familiar responses across two levels of confidence for these 40 participants. Once again, the data do not appear to follow a single, monotonically increasing function as old–new accuracy increases (the pattern that would be indicative of a single dimension). Instead, the function exhibits an apparent discontinuity moving from R_5 to F_6 (where old–new accuracy increases yet source accuracy decreases). As in Experiment 1, the deviation from a monotonically increasing function is in the direction predicted by the CDP model. However, once again, the deviation is not extreme, so these findings alone would not convincingly rule out a single-process account. Experiment 3 provides yet another test of this issue, which will help to establish the consistency of this result.

Sixteen participants in this experiment did not provide at least one response in each of the four relevant categories (R_5 , F_5 , R_6 , and F_6). Of those 16, six provided no R_5 responses, but they did provide F_5 , R_6 , and F_6 responses. Once again, this pattern is predicted by a threshold recollection model because, according to that model, recollection is not associated with degrees of confidence (whereas familiarity is). Alternatively, these participants may have simply had a high remember criterion, and the evidence is consistent with the latter interpretation. The average remember hit rate for these six participants was 15% (lower than the 27% remember hit rate of the 40 participants considered above), and their average remember false alarm rate was less than 1% (lower

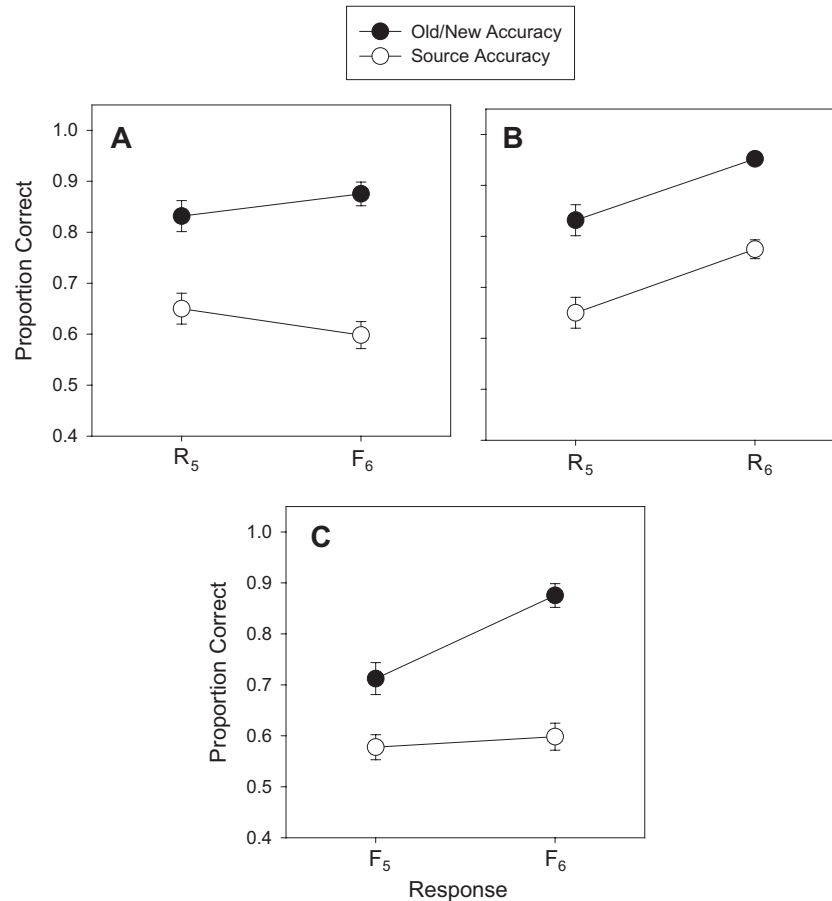


Figure 7. Proportion correct for old–new and source decisions in Experiment 2 for R₅ and F₆ responses (A), R₅ and R₆ responses (B), and F₅ and F₆ responses (C). R = remember; F = familiar.

than the 5% remember false alarm rate of the 40 participants considered above). In addition, their average source accuracy for R₆ responses was .72, and their average source accuracy for F₆ responses was .59. All six participants had scores of .50 or greater for F₆ responses. Thus, familiar judgments indicated a lesser degree of (not the absence of) recollection. These findings are consistent with the CDP model, given the assumption that the participants had a high criterion for making remember judgments.

Although the results from these six participants are easily reconciled with the CDP model, they are also consistent with a unidimensional memory strength model (one that need not assume more than one memory process). That is, across the three categories of responses (F₅, F₆, and R₆), source accuracy (55%, 59%, and 72%, respectively) tracked old–new accuracy (58%, 89%, and 99%, respectively). An additional five participants provided only recollection or only familiarity responses across the confidence scale, a pattern that also accords with a single-process memory strength account. Finally, the remaining five participants yielded results that did not help to distinguish between any of the models (e.g., they made very few ratings of 5 or 6, or they made R₆ responses almost exclusively because of an apparent ceiling effect).

Experiment 3

The results of Experiments 1 and 2 are consistent with the idea that recollection and familiarity are distinct, continuous processes, as envisioned by the CDP model. However, one way to explain these data while retaining a single-process, unidimensional interpretation would be to assume that participants did not follow instructions. Instead, they chose a remember response when they recollected criterial source details (location and/or color), and chose a familiar response when they recollected noncriterial source details (such as thoughts that occurred when the item was encoded). This might have happened because participants knew that a test of criterial source memory would immediately follow each remember–familiar judgment. If they were concerned about making an error on the criterial source test following a remember judgment when only noncriterial source details were recollected, they could avoid the issue by simply choosing one of the familiar options (in which case an error on the criterial source recollection test might be more forgivable in the eyes of the experimenter). Under this scenario, a rating of F₆ would not indicate a high-confidence familiarity-based decision but would instead indicate a high-confidence decision based on the recollection of noncriterial source details.

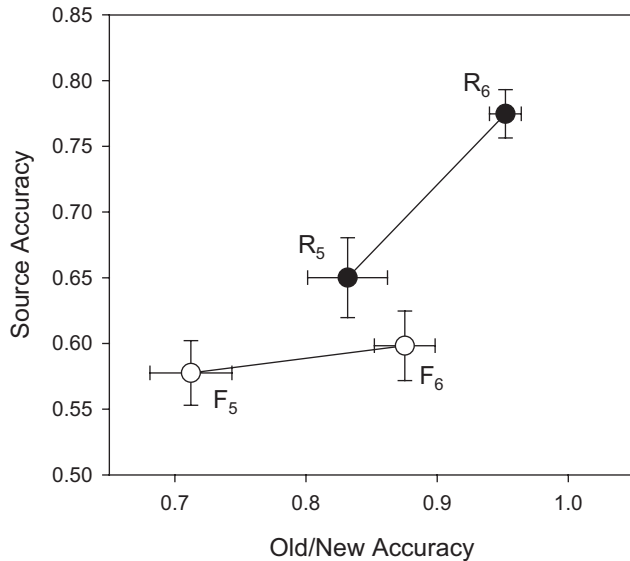


Figure 8. State-trace plot for Experiment 2 showing the mean old–new and source accuracies for F_5 , F_6 , R_5 , and R_6 . R = remember; F = familiar.

In Experiment 3, remember and familiar judgments were not followed by a source recollection test but were instead followed by the presentation of the next test item. In addition, the instructions at the beginning of the test session made no mention of a source recollection test, and the short practice session that preceded the experiment proper did not include a source recollection test. After the initial recognition test was completed, a surprise source recollection test was administered for all of the target items from the list. The advantage of administering a surprise source recall test at the end of the session is that it presumably eliminates any motivation that participants might have to cater their remember–familiar responses to the source test. A disadvantage is that the delayed source memory test resulted in a lower overall level of source accuracy, compared with the first two experiments, with quite a few participants failing to achieve above-chance performance (even for their R_{20} responses). However, we tested enough participants that a reasonably large group was obtained that (a) provided responses in all four of the relevant categories (so that the state-trace plot could be examined) and (b) achieved above-chance source accuracy, at least for their R_{20} responses.

Participants

Seventy-seven undergraduate students from UCSD participated in exchange for psychology course credit.

Materials and Design

These were all identical to Experiment 1.

Procedure

The procedure was the same as Experiment 1, except for the nature of the source information and the timing of the source memory test. During the study phase, each of the target words was paired with one of two questions (“Is it animate or inanimate?” or

“Would this fit in a shoebox?”). This was the source detail that participants would later be asked to recollect. Participants were unaware that they would be asked about source details at the end of the session, and the short practice session that preceded the experiment proper did not include a source memory test. After the initial recognition test was completed (during which confidence ratings and remember–familiar judgments were made), participants took a surprise source memory test. This test involved only the target items, which were re-randomized and presented one at a time for a source memory test. That is, for each target item, participants were asked to indicate the question that had originally appeared with the item.

Results

Because the source recollection test was delayed in this experiment, source recollection performance was noticeably lower than it was in Experiments 1 and 2, even when old–new confidence and accuracy were high. Indeed, for 27 of the 77 participants, source accuracy for R_{20} judgments was .54 or below. The average source accuracy for R_{20} responses over these 27 participants was only .48. We reasoned that including data from these participants (e.g., to test for source recollection differences for R_{16-19} responses vs. F_{20} responses) would only add random error to the analysis, so we limited our analysis to those who achieved source accuracy scores of at least .55 for their R_{20} responses. Also, because the main question was whether a state-trace plot similar to that observed in the first two experiments would be observed again here, we analyzed data from the remaining participants who had at least one rating of 16–19 and 20 for both remember and familiar responses. Thirty-three participants satisfied these criteria. The pattern of results we report below remains the same if the source accuracy inclusion criterion for R_{20} responses is set to .60, .70, or .80 (though fewer and fewer participants are included as the criterion is increased).

As before, the main comparison of interest concerned F_{20} responses versus R_{16-19} responses in the 33 participants who used the four relevant response categories. Figure 9A shows that old–new accuracy was higher for F_{20} responses (93%) than for R_{16-19} responses (86%), but source accuracy for R_{16-19} responses (63%) was higher than for F_{20} responses (59%). As in the first two experiments, the interaction between type of judgment (recollection vs. familiarity) and type of test (old–new vs. source) was significant, $F(1, 22) = 7.71, p = .011$.

Figure 9B shows the state-trace plot of the data. As was seen in Experiments 1 and 2, the plot appears to be nonmonotonic. The deviation from monotonicity is similar to what it was in the previous experiments and is in the same direction (which is the direction predicted by the CDP model). These results suggest that performance in those experiments was not appreciably affected by testing source recollection after each remember–familiar decision. That is, it seems likely that participants followed the instructions to provide remember responses when they believed that they were recollecting information about the item’s prior occurrence (whether they recollected criterial or noncriterial source details).

In all three experiments, old–new accuracy was higher for F_{20} (or, in Experiment 2, F_6) than for R_{16-19} (or R_5), but the reverse was true for source accuracy. Even so, source accuracy for R_{16-19} (or R_5) was not significantly higher than source accuracy for F_{20}

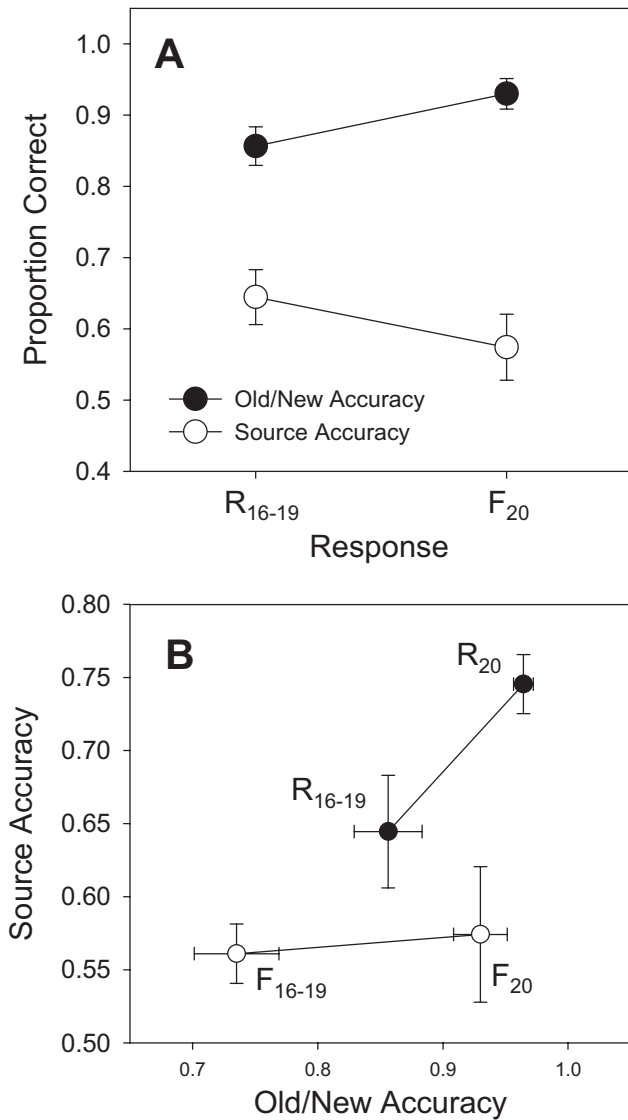


Figure 9. Proportion correct for old–new and source decisions in Experiment 3 for R₁₆₋₁₉ and F₂₀ responses (A), and state-trace plot showing the mean old–new and source accuracies for F₁₆₋₁₉, F₂₀, R₁₆₋₁₉, and R₂₀. R = remember; F = familiar.

(or F₆) in any of the three experiments. However, when participants who provided responses in all four response categories are combined across the three experiments to yield greater power (84 participants in all), the relevant differences are significant. That is, old–new accuracy was significantly higher for F₂₀–F₆ than for R₁₆₋₁₉–R₅, $t(83) = 3.23$, but source accuracy was significantly higher for R₁₆₋₁₉–R₅ than for F₂₀–F₆, $t(83) = 2.24$. This test provides an indication of the reliability of the apparent state-trace nonmonotonocities evident in Figures 5, 8, and 9B. In each case, the nonmonotonicity occurred at the same point. That is, moving from R₁₆₋₁₉ (or R₅) to F₂₀ (or F₆), a decrease, rather than an increase, was observed (a finding that is consistent with the two-dimensional CDP model). At least for the participants who supplied both remember and familiar judgments across multiple levels

of confidence (and, in Experiment 3, whose performance exhibited some capacity for source recollection), this test suggests that in the aggregate, the F₂₀–F₆ point is significantly farther to the right and significantly below that of the R₁₆₋₁₉–R₅ point in the state-trace plots.

General Discussion

The three experiments reported here used a variant of the remember–know procedure in which participants could directly indicate their level of confidence in remember–familiar decisions. Typically, remember judgments are made with high confidence and high accuracy, whereas know judgments are made with lower confidence and lower accuracy. This common pattern is consistent with dual-process theories that assume a threshold recollection process (such that the occurrence of recollection always yields high confidence) but is also consistent with a one-dimensional model based on signal-detection theory (according to which Remember and Know judgments reflect strong and weak memories, respectively). We found that when offered the explicit opportunity to do so, participants made remember judgments with varying levels of confidence and accuracy. This pattern has sometimes been reported in other studies as well (e.g., Rotello, Macmillan, & Reeder, 2004; Rotello et al., 2005), but the ability to recollect source details following low-confidence remember judgments has not been previously investigated. Thus, it could be (and has been) argued that inadequate instructions might have encouraged participants to use low-confidence remember judgments even when they did not recollect anything (e.g., Parks & Yonelinas, 2007). We used an objective test of source memory to investigate whether source recollection accuracy respects the remember–know distinction across different levels of confidence. The key finding was that remember judgments made with relatively low confidence and low old–new accuracy were consistently associated with higher source accuracy than familiar judgments made with higher confidence and higher old–new accuracy.

Our findings indicate that recollection can be weak, and familiarity can be strong. For that to be true, recollection and familiarity would have to be distinct, continuous memory variables as envisioned by the CDP model. Thus, our findings weigh against dual-process models that assume a threshold recollection process and against single-process models that do not allow for the possibility that recollection and familiarity contribute to item recognition memory. Our results also suggest that the unidimensional model of remember–know judgments proposed by Donaldson (1996), while providing a useful guide to the understanding of most remember–know studies, is incomplete. On average, remember–know judgments distinguish between strong and weak memories (in agreement with the standard unidimensional model), but when remember and know judgments are compared across different levels of confidence, our state-trace analyses suggests that performance is (or at least can be) based on two underlying dimensions.

Not all participants in our study distributed their remember–familiar responses across the rating scale, so it was not possible to include them in the state-trace analyses. Most of these participants made remember judgments with high confidence only (whereas familiar judgments were made with multiple levels of confidence). This pattern can be reconciled either with a dual-process model

that assumes threshold recollection (such as the HTSD model) or with a dual-process model that assumes continuous recollection (such as the CDP model) with a high criterion for making remember judgments. Other aspects of the data offer some support for the latter interpretation. Specifically, participants with this response profile had a lower remember hit rate and a lower remember false alarm rate than did the other participants, suggesting that they used a conservative criterion for making remember judgments. Whether or not the CDP model adequately explains the performance of these participants, it seems clear that it can offer a coherent explanation of the performance of the many participants who provided data that could be subjected to a state-trace analysis. For them, at least, it seems that recollection can be weak and familiarity can be strong (in which case recollection must be a continuous process).

If recollection is a continuous process, then it would not be possible to obtain valid estimates of recollection and familiarity by fitting the HTSD model to ROC data or by correcting remember and know judgments using the independence remember-know method. Both methods are widely used in cognitive neuroscience (e.g., Aggleton et al., 2005; Yonelinas, Kroll, Dobbins, Lazzara, & Knight, 1998), but both rely on the assumption that recollection is a threshold process. In addition, it would not be possible to use simple remember-know judgments to effectively differentiate between recollection and familiarity in neuroimaging studies (e.g., Eldridge, Knowlton, Furmanski, Bookheimer, & Engel, 2000). Even if the CDP model is correct, Wixted and Mickes (2010) showed that it would still be the case that on average, remember judgments reflect higher confidence and higher accuracy than do know judgments (which the unidimensional signal-detection model assumes as well). As such, any difference in neural activity associated with remember judgments and know judgments could reflect a difference between strong and weak memories, not between recollection and familiarity. To effectively investigate recollection and familiarity using the remember-know procedure, methods that avoid this strength confound, such as the method we used here, are needed (cf. Kirwan, Wixted, & Squire, 2008; Wais, Squire, & Wixted, 2010).

Our findings do not necessarily indicate that recollection and familiarity always play a role in item recognition memory. For example, all three of our experiments included conspicuous source attributes, and it is conceivable that participants attempted to recollect those attributes during the recognition test in a way that they might not have, had no explicit source attributes accompanied the words at study. In addition, participants in our study were asked to make remember-familiar judgments and that may also have induced them to rely on recollection more than they ordinarily would. Still, at least under the conditions we used, it seems clear that participants can rely on continuous recollection and familiarity signals, and when they do, they can provide valid information about which memory process mainly informed their recognition decisions.

Single-Process Versus Dual-Process Terminology

The one-dimensional UVSD model of recognition memory is easily reconciled with a single-process model (e.g., one that assumes that all decisions are based on familiarity), and this may explain why the unidimensional model is often thought to be an

inherently single-process account (e.g., Malmberg, 2008). However, the CDP model reduces to the UVSD model when the task involves making an old-new decision by assuming that recollection and familiarity are aggregated into a single memory strength dimension. Thus, the mere fact that decisions are assumed to be based on a single-dimension of memory strength does not necessarily imply that only one memory process is involved. Absent compelling findings to the contrary, parsimony would demand that a unidimensional model be regarded as a single-process model, but our argument is that the findings that we present here provide evidence for the existence of two continuous processes. Because both processes are continuous, they could be aggregated on an old-new task in such a way that the standard one-dimensional UVSD model would still apply.

When should a theory be regarded as a dual-process theory? Dual-process theory is sometimes equated with the notion that one memory signal is dichotomous and the other is continuous (e.g., Heathcote, Raymond, & Dunn, 2006; Hautus, Macmillan, & Rotello, 2008). If the scalar properties of memory signals differ in that way, it seems reasonable to suppose that they arise from different processes. The Yonelinas (1994) HTSD model is the cardinal example of such a model, with recollection construed as a dichotomous process and familiarity construed as a continuous process. However, from our point of view, the distinction between dichotomous and continuous processes is not related to the distinction between recollection and familiarity. Familiarity, whether construed as a dichotomous or a continuous process, consists of a memory signal associated with the item itself, whereas recollection, whether construed as a dichotomous or a continuous process, consists of the retrieval of associated (e.g., source) information. That is, the essential difference between the subjective experience of recollection and familiarity concerns the content of the memory signal, not the scalar properties of the memory signal. Familiarity is usually considered a continuous process, as in the HTSD model, but in multinomial models it is sometimes regarded as a dichotomous process (e.g., Buchner, Erdfelder, & Vaterrodt-Plünnecke, 1995). Recollection is sometimes considered a dichotomous process, as in the HTSD model, but it is regarded as a continuous process in the CDP model as well as in other models (e.g., Rotello, Macmillan, Reeder, & Wong, 2004). Thus, in our view, the distinction between dichotomous and continuous processes is orthogonal to the distinction between recollection and familiarity.

These terminological considerations help to clarify the relationship between the CDP model and a previous model proposed by Heathcote et al. (2006). They investigated whether source memory is based on a dichotomous or continuous variable using ROC analysis. One representative experiment involved a list discrimination procedure in which participants first studied two lists of words. On a subsequent recognition test, they were presented with items from each list and with new items (what they called the *with new* condition). Participants were instructed to classify a test item as old if it appeared in one of the lists (e.g., List 1) and to classify it as new if it appeared in the other list (e.g., List 2) or if it was a new word. Thus, the task involved elements of both an old-new recognition memory test and a source memory test. The words presented on Lists 1 and 2 were presumably equally familiar, so the ability to discriminate list membership was assumed to be based on retrieved source information, not item familiarity. The retrieval of source information is sometimes thought to involve a

dichotomous recollection process, but Heathcote et al. (2006) reported clear evidence that it was instead based on a continuous variable that they termed *specific strength*. Their model assumes that in addition to specific strength, item familiarity is also taken into account (and is, in fact, combined in additive fashion with specific strength) because it helps to distinguish items that appeared on one of the lists from new items. That is, for the *with new* condition, they suggested that “decisions are based on the sum of specific strength, $s_T(I)$, and familiarity, $F(I)$ ” (p. 513). The idea that specific strength and familiarity are combined into a singular memory strength variable on this task corresponds closely to the idea that recollection and familiarity are combined into a singular memory strength variable when making an ordinary old–new recognition decision (as the CDP model assumes).

Although the model proposed by Heathcote et al. (2006) is similar to the CDP model in several important respects, the terminology they used appears to suggest otherwise. For example, what they refer to as specific strength is what we refer to as continuous recollection (they reserved the term *recollection* for dichotomous recollection). In our view, the term *recollection* should not be reserved for a dichotomous process and should instead be used to refer to the retrieval of information that is associated with item information (i.e., information that is not represented in the copy cue). Thus, specific strength might be better termed continuous recollection because the retrieval of such information gives rise to the subjective sense of recollection (and to remember judgments). In addition, although the model proposed by Heathcote et al. (2006) distinguishes between two dimensions of memory strength—specific strength and item familiarity—they nevertheless characterized it as a single-process model because both sources of evidence are assumed to be continuous and are derived from different cue-matching operations. Heathcote et al. (2006) acknowledged that referring to their account as a single-process model may be a misnomer, and we agree. In our view, it seems simpler to think of their model as a dual-process model that involves a continuous recollection signal (what they call specific strength) and a continuous familiarity signal. Terminological differences notwithstanding, both models assume that old–new recognition can be based on the combination of two continuous memory signals: item familiarity and source memory. Heathcote et al.’s (2006) model does not necessarily assume that specific strength and familiarity are additively combined in ordinary old–new recognition (i.e., one that does not involve an explicit source memory component), whereas the CDP model does make that assumption.

The CDP Model Versus Multidimensional Models of Source Memory

A number of other signal-detection models have also drawn a distinction between a continuous memory signal for old–new memory decisions and a continuous memory signal for source memory decisions (Banks, 2000; DeCarlo, 2003; Hautus et al., 2008). These models apply to tasks in which participants are asked to make an old–new decision and a source memory decision separately. Figure 10 illustrates the two dimensions of a multidimensional source memory model (from Hautus et al., 2008), and it also illustrates a positive correlation between old–new memory strength and source memory strength. The positive correlation

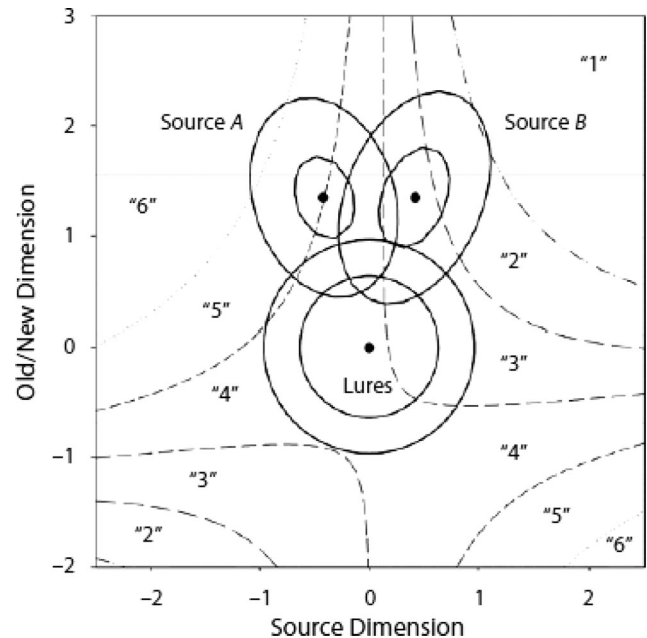


Figure 10. Two-dimensional decision space for recognition and source memory (with likelihood ratio decision bounds) proposed by Hautus et al. (2008). The numbers represent source confidence ratings on a 6-point scale (6 = *sure Source A*; 1 = *sure Source B*).

reflects the fact that items that receive higher confidence ratings on one task also tend to receive higher confidence ratings on the other.

How do these two dimensions (old–new strength and source strength) relate to the three dimensions of the CDP model (aggregate memory strength, recollection strength, and familiarity strength)? As indicated earlier, familiarity consists of a memory signal that corresponds to the test item itself (i.e., that corresponds to the copy cue), whereas recollection consists of the retrieval of information beyond the test item. By that simple definition, source memory is an example of continuous recollection. Thus, the source memory axis of a multidimensional signal-detection model is similar to the recollection axis of the CDP model. They differ only in that the source memory axis usually represents the strength of a *critical* source memory signal, whereas the recollection axis of the CDP model represents the strength of any information that is recollected (including, but not limited to, criterial source recollection). Although only criterial source recollection can contribute to a criterial source memory decision (e.g., did this item appear at the top of the screen or the bottom of the screen?), which is what multidimensional source memory models attempt to explain, both criterial recollection and noncriterial recollection can contribute to old–new recognition, which is what the CDP model attempts to explain.

The old–new memory strength axis of some multidimensional signal-detection models corresponds to the familiarity axis of the CDP model. In DeCarlo’s (2003) multidimensional model, for example, the old–new axis was referred to as *familiarity*, as if old–new decisions are based on familiarity but source decisions are based on an altogether different memory signal (continuous source recollection). This model accounts for the positive correlation between old–new memory strength and source memory

strength by assuming that some items receive more attention (or are more deeply processed) at study than do other items. For more deeply processed items, familiarity and source recollection would both be higher than for items processed in a more shallow fashion (hence, the positive correlation). DeCarlo's (2003) model differs from the CDP model in that it does not assume an aggregate memory strength axis and, therefore, does not assume that source recollection contributes to old–new decisions. Instead, it assumes that old–new decisions are based on familiarity and that source decisions are based on what we would call continuous recollection. Hautus et al. (2008) did not specifically state whether the old–new axis in their multidimensional model reflects only item familiarity or whether it reflects a combination of item familiarity and source information. They did briefly suggest that some of the same information contributes to item and source decisions, which at least raises the possibility that old–new strength consists of an aggregated memory signal (similar to the old–new axis in the CDP model). Either way, it seems clear that an interesting question for future studies to address is whether the retrieval of source information contributes in any way to old–new decisions (as the CDP model assumes it does) or whether such decisions are based solely on item information (i.e., solely on familiarity).

A recent multidimensional source memory model proposed by Onyper, Zhang, and Howard (2010) is more explicit about assuming continuous source recollection and item familiarity signals that can each be associated with a range of confidence. However, for old–new decisions, this model retains the flavor of a high-threshold account (such as the HTSD model) by assuming that a purely recollection-based decision is made whenever any degree of recollection occurs (though it is not clear why recollection would ever yield low confidence if lures can never achieve that state). According to this model, a familiarity-based decision is made only if recollection fails. Thus, the model does not assume that the continuous recollection and familiarity signals are aggregated to make an old–new decision, as the CDP model does. Instead, as with the Yonelinas (1994) HTSD model, individual old–new decisions are assumed to be based either on recollection or on familiarity (never both together). Moreover, their model does not assume that recollection is a continuous signal-detection process because it assumes that lures do not generate a continuous recollection signal, even though targets do.

Conclusion

Despite some similarities discussed above, there is one key difference between the CDP model and prior multidimensional models of memory. The CDP model is a model of remember–know judgments; multidimensional source memory models are not. Even if a multidimensional source memory model assumes that old–new recognition memory is supported by two continuous memory signals (such as source recollection and item familiarity), it does not necessarily follow that remember–know judgments effectively distinguish between old–new decisions that are based mainly on recollection and old–new decisions that are based mainly on familiarity. Whether they do is a further empirical question, and much prior work suggests that as ordinarily used, remember–know judgments merely reflect different degrees of overall memory strength (and, therefore, do not convincingly distinguish between recollection and familiarity). However, our

current findings suggest that once steps are taken to eliminate the typical strength confound, remember–know judgments can be used to effectively distinguish between decisions that involve different degrees of recollection and familiarity.

Finally, it seems important to emphasize that our results only point to the existence of two continuous dimensions of information that differ in terms of the content of the mnemonic signal. One dimension reflects the degree of source information retrieved (recollection), whereas the other reflects a context-free sense of prior occurrence (familiarity). From one perspective, which is common in experimental psychology, these dimensions should only be viewed as separate processes if they are differentially affected by experimental manipulations (such as a levels-of-processing manipulation). From another perspective, which is more common in cognitive neuroscience, these dimensions should only be viewed as separate processes if they are subserved by different brain structures that rely on different neurocomputational algorithms. Our findings do not establish the existence of separate processes in either sense. Instead, our findings point to the existence of separate mnemonic signals that are compatible with the longstanding distinction between recollection and familiarity.

References

- Aggleton, J. P., Vann, S. D., Denby, C., Dix, S., Mayes, A. R., Roberts, N., & Yonelinas, A. P. (2005). Sparing of the familiarity component of recognition memory in a patient with hippocampal pathology. *Neuropsychologia*, *43*, 1810–1823. doi:10.1016/j.neuropsychologia.2005.01.019
- Atkinson, R. C., & Juola, J. F. (1973). Factors influencing the speed and accuracy of word recognition. In S. Kornblum (Ed.), *Attention and performance IV* (pp. 583–612). New York, NY: Academic Press.
- Bamber, D. (1979). State-trace analysis: A method of testing simple theories of causation. *Journal of Mathematical Psychology*, *19*, 137–181. doi:10.1016/0022-2496(79)90016-6
- Banks, W. P. (2000). Recognition and source memory as multivariate decision processes. *Psychological Science*, *11*, 267–273. doi:10.1111/1467-9280.00254
- Buchner, A., Erdfelder, E., & Vaterrodt-Plünnecke, B. (1995). Toward unbiased measurement of conscious and unconscious memory processes within the process dissociation framework. *Journal of Experimental Psychology: General*, *124*, 137–160. doi:10.1037/0096-3445.124.2.137
- Coltheart, M. (1981). The MRC psycholinguistic database. *Quarterly Journal of Experimental Psychology*, *33A*, 497–505.
- DeCarlo, L. T. (2003). An application of signal detection theory with finite mixture distributions to source discrimination. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 767–778. doi:10.1037/0278-7393.29.5.767
- Donaldson, W. (1996). The role of decision processes in remembering and knowing. *Memory & Cognition*, *24*, 523–533. doi:10.3758/BF03200940
- Dunn, J. C. (2004). Remember–know: A matter of confidence. *Psychological Review*, *111*, 524–542. doi:10.1037/0033-295X.111.2.524
- Dunn, J. C. (2008). The dimensionality of the remember–know task: A state-trace analysis. *Psychological Review*, *115*, 426–446. doi:10.1037/0033-295X.115.2.426
- Egan, J. P. (1958). *Recognition memory and the operating characteristic* (Technical Note AFCRC-N-58–51). Bloomington, IN: Indiana University, Hearing and Communication Laboratory.
- 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. doi:10.1038/80671

- Gardiner, J. M. (1988). Functional aspects of recollective experience. *Memory & Cognition*, *16*, 309–313. doi:10.3758/BF03197041
- Gardiner, J., Richardson-Klavehn, A., & Ramponi, C. (1997). On reporting recollective experiences and “direct access to memory systems.” *Psychological Science*, *8*, 391–394. doi:10.1111/j.1467-9280.1997.tb00431.x
- Hautus, M. J., Macmillan, N. A., & Rotello, C. M. (2008). Toward a complete decision model of item and source recognition. *Psychonomic Bulletin & Review*, *15*, 889–905. doi:10.3758/PBR.15.5.889
- Heathcote, A. (2003). Item recognition memory and the ROC. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 1210–1230. doi:10.1037/0278-7393.29.6.1210
- Heathcote, A., Raymond, F., & Dunn, J. (2006). Recollection and familiarity in recognition memory: Evidence from ROC curves. *Journal of Memory and Language*, *55*, 495–514. doi:10.1016/j.jml.2006.07.001
- Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language*, *30*, 513–541. doi:10.1016/0749-596X(91)90025-F
- Jang, Y., Wixted, J. T., & Huber, D. E. (2009). Testing signal-detection models of yes/no and two-alternative forced-choice recognition memory. *Journal of Experimental Psychology: General*, *138*, 291–306. doi:10.1037/a0015525
- Kirwan, C. B., Wixted, J. T., & Squire, L. R. (2008). Activity in the medial temporal lobe predicts memory strength, whereas activity in the prefrontal cortex predicts recollection. *The Journal of Neuroscience*, *28*, 10541–10548. doi:10.1523/JNEUROSCI.3456-08.2008
- Malmberg, K. J. (2008). Recognition memory: A review of the critical findings and an integrated theory for relating them. *Cognitive Psychology*, *57*, 335–384. doi:10.1016/j.cogpsych.2008.02.004
- Mandler, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological Review*, *87*, 252–271. doi:10.1037/0033-295X.87.3.252
- McCabe, D. P., & Geraci, L. D. (2009). The influence of instructions and terminology on the accuracy of remember-know judgments. *Consciousness and Cognition*, *18*, 401–413. doi:10.1016/j.concog.2009.02.010
- Mickes, L., Wais, P. E., & Wixted, J. T. (2009). Recollection is a continuous process: Implications for dual process theories of recognition memory. *Psychological Science*, *20*, 509–515. doi:10.1111/j.1467-9280.2009.02324.x
- Montaldi, D., Spencer, T. J., Roberts, N., & Mayes, A. R. (2006). The neural system that mediates familiarity memory. *Hippocampus*, *16*, 504–520. doi:10.1002/hipo.20178
- Newell, B. R., & Dunn, J. C. (2008). Dimensions in data: Testing psychological models using state-trace analysis. *Trends in Cognitive Sciences*, *12*, 285–290. doi:10.1016/j.tics.2008.04.009
- Onyper, S. V., Zhang, Y. X., & Howard, M. W. (2010). Some-or-none recollection: Evidence from item and source memory. *Journal of Experimental Psychology: General*, *139*, 341–364. doi:10.1037/a0018926
- Parks, C. M., & Yonelinas, A. P. (2007). Moving beyond pure signal detection models: Comment on Wixted (2007). *Psychological Review*, *114*, 188–201. doi:10.1037/0033-295X.114.1.188
- Rotello, C. M., Macmillan, N. A., Hicks, J. L., & Hautus, M. J. (2006). Interpreting the effects of response bias on remember-know judgments using signal detection and threshold models. *Memory & Cognition*, *34*, 1598–1614. doi:10.3758/BF03195923
- Rotello, C. M., Macmillan, N. A., & Reeder, J. A. (2004). Sum-difference theory of remembering and knowing: A two-dimensional signal detection model. *Psychological Review*, *111*, 588–616. doi:10.1037/0033-295X.111.3.588
- Rotello, C. M., Macmillan, N. A., Reeder, J. A., & Wong, M. (2005). The remember response: Subject to bias, graded, and not a process-pure indicator of recollection. *Psychonomic Bulletin & Review*, *12*, 865–873. doi:10.3758/BF03196778
- Rotello, C. M., & Zeng, M. (2008). Analysis of RT distributions in the remember-know paradigm. *Psychonomic Bulletin & Review*, *15*, 825–832. doi:10.3758/PBR.15.4.825
- Starns, J. J., & Ratcliff, R. (2008). Two dimensions are not better than one: STREAK and the univariate signal detection model of remember/know performance. *Journal of Memory and Language*, *59*, 169–182. doi:10.1016/j.jml.2008.04.003
- Wais, P. E., Mickes, L., & Wixted, J. T. (2008). Remember/know judgments probe degrees of recollection. *Journal of Cognitive Neuroscience*, *20*, 400–405. doi:10.1162/jocn.2008.20041
- Wais, P. E., Squire, L. R., & Wixted, J. T. (2010). In search of recollection and familiarity signals in the hippocampus. *Journal of Cognitive Neuroscience*, *22*, 109–123. doi:10.1162/jocn.2009.21190
- Wixted, J. T. (2007). Dual-process theory and signal-detection theory of recognition memory. *Psychological Review*, *114*, 152–176. doi:10.1037/0033-295X.114.1.152
- Wixted, J. T., & Mickes, L. (2010). A continuous dual-process model of remember/know judgments. *Psychological Review*, *117*, 1025–1054.
- Wixted, J. T., & Stretch, V. (2004). In defense of the signal-detection interpretation of remember/know judgments. *Psychonomic Bulletin & Review*, *11*, 616–641. doi:10.3758/BF03196616
- 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. doi:10.1037/0278-7393.20.6.1341
- Yonelinas, A. P. (2001). Consciousness, control, and confidence: The 3 Cs of recognition memory. *Journal of Experimental Psychology: General*, *130*, 361–379. doi:10.1037/0096-3445.130.3.361
- Yonelinas, A. P., Kroll, N. E., Dobbins, I., Lazzara, M., & Knight, R. T. (1998). Recollection and familiarity deficits in amnesia: Convergence of remember-know, process dissociation, and receiver operating characteristic data. *Neuropsychology*, *12*, 323–339. doi:10.1037/0894-4105.12.3.323

Received December 21, 2010

Revision received May 15, 2011

Accepted June 17, 2011 ■