

# A Continuous Dual-Process Model of Remember/Know Judgments

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

The dual-process theory of recognition memory holds that recognition decisions can be based on recollection or familiarity, and the remember/know procedure is widely used to investigate those 2 processes. Dual-process theory in general and the remember/know procedure in particular have been challenged by an alternative strength-based interpretation based on signal-detection theory, which holds that remember judgments simply reflect stronger memories than do know judgments. Although supported by a considerable body of research, the signal-detection account is difficult to reconcile with G. Mandler's (1980) classic "butcher-on-the-bus" phenomenon (i.e., strong, familiarity-based recognition). In this article, a new signal-detection model is proposed that does not deny either the validity of dual-process theory or the possibility that remember/know judgments can—when used in the right way—help to distinguish between memories that are largely recollection based from those that are largely familiarity based. It does, however, agree with all prior signal-detection-based critiques of the remember/know procedure, which hold that, as it is ordinarily used, the procedure mainly distinguishes strong memories from weak memories (not recollection from familiarity).

*Keywords:* recognition, memory, familiarity, recollection, signal detection theory

The strength of memory and the content of memory are, in theory, separable and independent properties, but they often have been viewed as fundamentally incompatible ideas. In this article, we argue that the attempt to understand memory in terms of either strength or content is misplaced because both ideas are needed. Nowhere is this more evident than in the ongoing debate over dual-process versus single-process interpretations of remember/know judgments.

The concept of memory strength applies naturally to a variety of behavioral measures that tend to covary, such as confidence, accuracy, and reaction time. Generally speaking, memories are said to be strong when they are associated with relatively high confidence, high accuracy, and fast reaction times. The idea that memories vary in strength is also usefully conceptualized in terms of signal-detection theory. The standard unequal-variance version of this theory holds that targets and lures on a recognition memory test have normally distributed memory strengths, with the mean and variance of the former being greater than the mean and variance of the latter. If the strength of memory associated with a particular test item exceeds a decision criterion, then that item is declared to be old; otherwise, it is declared to be new. Conceptualizing recognition memory along these lines helps to make sense of the way in which false alarm rates change across conditions, as

in the mirror effect (Glanzer, Adams, Iverson, & Kim, 1993), and it helps to make sense of the relationship between hit rates and false alarm rates as a function of confidence, as in analyses of the receiver operating characteristic (ROC; Egan, 1958). Figure 1 illustrates the standard unequal-variance signal-detection model that has guided thinking about recognition memory for more than 50 years.

Despite its utility in some domains, the concept of memory strength seems like a woefully inadequate construct to capture the richness of memory. Anderson and Bower (1972), in an interesting discussion of this issue, had this to say:

The important feature to note about the strength theory of recognition is that it is "ahistorical"; that is, it assumes that a subject makes recognition decisions about an item not on the basis of detailed memory of the past history of occurrences of the item, but rather on the basis of a single measure which reflects to some extent its past frequency, recency, and duration of exposure. It is this ahistorical character of strength theory which is the source of all its weaknesses. (p. 98)

They went on to note how the notion of memory strength is compatible with signal-detection models and that the memory-strength variable in such models is often named "familiarity." The basic problem with this notion was summarized by Anderson and Bower (1972) in the following way:

However, the evidence is now available that an undifferentiated strength of familiarity concept is not sufficiently rich to account for the subject's ability to differentiate sets of items. (p. 100)

In their view, and in the view of many others, memory is much more elaborate than an undifferentiated strength variable could ever be. To take one example used by Anderson and Bower (1972), when participants study two lists in succession, they can, if asked to do so, selectively recall (or selectively recognize) items

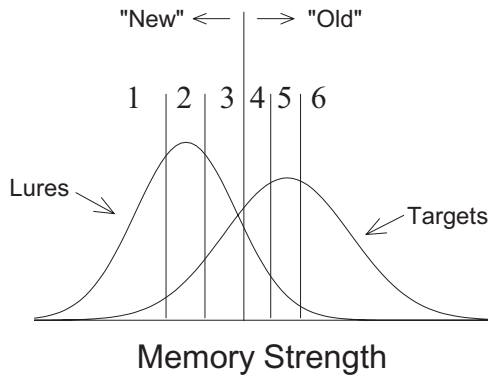
---

This article was published Online First September 13, 2010.

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

This work was supported by National Institute of Mental Health Grant R01MH082892 to John T. Wixted. 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.

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



*Figure 1.* An illustration of the unequal-variance model of recognition memory. The distribution of memory strengths for targets has a higher mean and greater variance than the distribution of memory strengths for lures. Confidence ratings made on a 6-point scale are based on five criteria arrayed along the memory-strength axis.

from the first list even though items from the second list, having been presented more recently, are presumably stronger than items from the first.

The notion of an ahistorical memory-strength variable fits quite naturally with the view of familiarity that was prevalent in the 1960s, but it also fits with the view of familiarity that is prevalent today as one of the two components of the dual-process theory of recognition memory. Contemporary dual-process theory holds that successful recognition can be based on a sense of familiarity generated by the test item or on the recollection of source details associated with the test item. In these theories, familiarity is generally viewed as a continuous variable, whereas recollection is usually thought to be an either/or categorical variable. In other words, familiarity comes in degrees ranging from low to high, but recollection either does or does not occur (Atkinson & Juola, 1973, 1974; Jacoby, 1991; Mandler, 1980; Yonelinas, 1994). Unlike recollection, familiarity is thought to be an undifferentiated, ahistorical memory-strength variable. This idea is the legacy of the memory-strength theories in the 1960s that Anderson and Bower (1972) criticized as being inappropriately ahistorical, but dual-process theory escapes that criticism because the historical information that participants clearly have access to is captured by the categorical recollection process.

In recent years, advocates of signal-detection theory have argued often against the dual-process view of recognition memory and have argued instead that recognition decisions are best construed as being based on a singular memory-strength variable, not on two separate processes (e.g., Donaldson, 1996; Dougal & Rotello, 2007; Dunn, 2004, 2008; Hirshman & Henzler, 1998; Slotnick & Dodson, 2005; Shimamura & Wickens, 2009). However, modern-day signal-detection theorists appear to be less wedded to the idea that memory strength is properly conceptualized as an ahistorical familiarity variable. Instead, in the view of some, memory strength is better characterized as the amount of mnemonic information that is retrieved. Anderson and Bower (1972) initiated this way of thinking by proposing that memory strength is best construed as an “evidence” variable that consists of the amount of source information retrieved about the test item. Similarly, Dunn (2008) labeled the memory-strength axis of signal-

detection theory as “strength of evidence.” Slotnick and Dodson (2005) advocated a single-process signal-detection model and showed that as confidence in the old/new decision increases, the amount of available source recollection increases as well. Similarly, DeCarlo (2003) and Hautus, Macmillan, and Rotello (2008) proposed signal-detection models in which old/new memory strength was correlated with source-recollective strength (as if the degree of memory strength in an old/new recognition task is partly a function of the degree of source recollection). In a new signal-detection model, Shimamura and Wickens (2009) argued that memory strength is determined by hierarchical relational binding of feature units, a concept that is closer to what dual-process theorists construe as recollection than to familiarity. Indeed, if one specifically named the memory-strength dimension envisioned by these various signal-detection theorists in terms used by dual-process theorists (which is ordinarily not done), *recollection* would probably be a more applicable label than *familiarity*. This recollection-like signal differs from the recollection process envisioned by traditional dual-process accounts in that it is continuous rather than categorical, but it is similar in that it is thought to contain historical information. In this view, the strength of memory and the content of memory are not incompatible concepts; instead, they are essentially one and the same (because strength is construed as the amount of content retrieved).

Whether the memory-strength variable in signal-detection theory is conceptualized as being akin to an ahistorical continuous familiarity signal or to a continuous recollection-like signal, the point is that recognition decisions in signal-detection models are usually thought to be based on one process, not two. The disagreement between the strength-based signal-detection point of view and the dual process point of view is particularly evident in the debate over the interpretation of remember/know judgments, which have often been used to study recollection and familiarity (see Gardiner & Richardson-Klavehn, 2000, for a brief review). We argue (as many others have) that the concept of memory strength is essential to understanding the basis of remember/know judgments. However, we also introduce a new dual-process theory of remember/know judgments, one that respects signal-detection theory (because the data seem to demand it) but that also assumes that recognition decisions involve recollection and familiarity and that participants have some capacity to accurately indicate which process mainly informed their recognition decision. No prior signal-detection account allows for the latter possibility, and no prior dual-process account respects signal-detection theory. Given the remarkably widespread (and still growing) use of the remember/know procedure in a variety of fields, it seems important to develop an adequate theory of exactly what remember and know judgments represent.

### Dual-Process Theory and the Remember/Know Procedure

Tulving (1985) distinguished between two states of awareness associated with the conscious experience of memory. One state, corresponding to retrieval from episodic memory, involves the awareness of a past event as being autobiographical in nature. Another state, corresponding to retrieval from semantic memory, involves the awareness of previously acquired knowledge but without any autobiographical component. Tulving (1985) pro-

posed that participants could indicate which state of awareness applied to a particular memory by saying “remember” if it was retrieved from episodic memory or “know” if it was retrieved from semantic memory.

In subsequent years, the use of the remember/know procedure evolved and became more commonly used in relation to the dual-process theory of recognition memory. A great deal of experimental evidence has been offered in support of the idea that recognition memory involves two processes (see Mandler, 1980, and Yonelinas, 2002, for detailed reviews). In addition to experimental evidence bearing on the issue, the distinction between recollection and familiarity is also suggested by the almost universal experience of encountering the familiar face of an individual who you cannot specifically remember having met before. In Mandler’s (1980) classic “butcher-on-the-bus” anecdote, the experience involves encountering a man on a bus whose face is so familiar that it compels a search of memory to determine the source of that feeling. The sense of familiarity is strong enough to elicit high confidence that the man was previously encountered, even though no information has yet been recollected about who he is or where the prior encounter might have occurred. Ultimately, the search of memory may be successful and yield the confirming recollection of context (“That’s the butcher from the supermarket!”). This anecdote not only illustrates the separate recollection and familiarity processes, it also suggests that people can experience strong memory based on an ahistorical familiarity process and that they are able to consciously appreciate that fact.

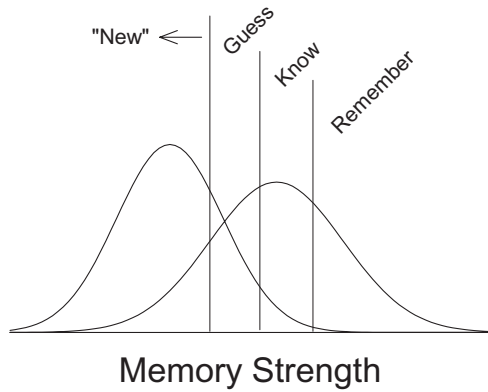
If people do appreciate when recollection and familiarity occur, it seems reasonable to suppose that they can also verbally report which process supported a given recognition decision, and this is how the remember/know procedure has been used in recent years. More specifically, in most studies involving the remember/know procedure, participants are instructed to indicate directly, for each “old” decision they make, whether it was based on recollection or familiarity by saying “remember” or “know,” respectively. This method and variants of it are ever more widely used in the fields of experimental psychology, cognitive neuroscience, and experimental psychopathology to study these two memory processes. Indeed, a search of the literature suggests that the procedure has never been more popular, having been used far more since 2006 than in any comparable period of time previously. Table 1 lists the remember/know studies published in just 2009, a year in which the remember/know procedure was used more than in any previous year, and the journal names illustrate the diversity of fields that have embraced this approach. In nearly all of these studies, remember judgments are assumed to reflect recollection, and know judgments are assumed to reflect familiarity.

The idea that remember and know judgments are based on recollection and familiarity has been strongly challenged by an alternative memory-strength interpretation that is grounded in signal-detection theory (Donaldson, 1996). According to this idea, remember and know judgments reflect different degrees of memory strength, not qualitatively different memory processes. The strength-based signal-detection interpretation of remember/know judgments is illustrated in Figure 2. This illustration applies to a common variant of the original procedure in which participants are asked to make a remember/know/guess judgment for each item that is declared to be old.

Table 1  
*Articles Published in 2009 Using the Remember/Know Procedure*

Author(s)	Journal
Gomez, Rousset, and Baciu	<i>Acta Psychologica</i>
Dewhurst, Conway, and Brandt	<i>Applied Cognitive Psychology</i>
Harkin and Kessler	<i>Behaviour Research and Therapy</i>
Hudon, Belleville, and Gauthier	<i>Brain and Cognition</i>
Parker, Buckley, and Dagnall (2008)	<i>Brain and Cognition</i>
Friedman, de Chastelaine, Nessler, and Malcolm	<i>Brain Research</i>
Geraci, McCabe, and Guillory	<i>Consciousness and Cognition: An International Journal</i>
McCabe and Geraci	<i>Consciousness and Cognition: An International Journal</i>
Klumpp, Amir, and Garfinkel	<i>Depression and Anxiety</i>
Dorfel, Wener, Schaefer, von Kummer, and Karl	<i>European Journal of Neuroscience</i>
Viskontas, Carr, Engel, and Knowlton	<i>Hippocampus</i>
Vilberg and Rugg	<i>Human Brain Mapping</i>
Carr, Viscontas, Engel, and Knowlton	<i>Journal of Cognitive Neuroscience</i>
Sauerland and Sporer	<i>Journal of Experimental Psychology: Applied</i>
Aizpurua, Garcia-Bajos, and Migueles	<i>Journal of General Psychology</i>
Jermann, Van der Linden, Laurencon, and Schmitt	<i>Journal of Psychopathology and Behavioral Assessment</i>
Clarys, Bugaiska, Tapia, and Baudouin	<i>Memory</i>
Lemogne, Bergouignan, Piolino, Jouvent, Allilaire, and Fossati	<i>Memory</i>
MacLaverly and Hertzog	<i>Memory</i>
Parker and Dagnall	<i>Memory</i>
McCabe and Geraci	<i>Memory &amp; Cognition</i>
Voss and Paller	<i>NeuroImage</i>
Johnson, McDuff, Rugg, and Norman	<i>Neuron</i>
Tanweer, Rathbone, and Souchay	<i>Neuropsychologia</i>
Vilberg and Rugg	<i>NeuroReport</i>
Daury	<i>Perceptual and Motor Skills</i>
Cohn, Moscovitch, Lahat, and McAndrews	<i>Proceedings of the National Academy of Sciences USA</i>
Stoetinger, Kaiser, and Perner	<i>Psychological Research</i>
Skinner and Fernandes	<i>Psychology and Aging</i>
de Chastelaine, Friedman, Cycowicz, and Horton	<i>Psychophysiology</i>
Peker and Tekcan	<i>Social Psychology</i>
Hess, Emery, and Queen	<i>The Journal of Gerontology</i>
Migo, Montaldi, Norman, Quamme, and Mayes	<i>The Quarterly Journal of Experimental Psychology</i>

According to this interpretation, a remember judgment is made when the strength of the memory signal exceeds a high criterion, in much the same way that a high-confidence old decision is made when the memory-strength signal exceeds a high criterion (see Figure 1). By contrast, a know judgment is made when the strength of the memory signal only exceeds the next lower criterion. In this case, the signal is strong enough for the item to be declared old, but it is not strong enough to declare that it is remembered (so it is declared to be known instead). A still lower criterion is used to make a guess judgment. This interpretation can be applied whether



*Figure 2.* The signal-detection interpretation of remember/know/guess judgments proposed by Donaldson (1996). If the memory strength of a test item does not exceed the leftmost criterion, it is declared to be “new.” If it does exceed the leftmost criterion, it is declared to be “old” and is further classified as a guess, know, or remember judgment, depending on whether or not it exceeds a higher criterion as well. Remember judgments are made when memory strength exceeds the highest criterion, know judgments are made when memory strength only exceeds the middle criterion, and guess judgments are made when memory strength only exceeds the leftmost criterion.

one regards the singular memory-strength variable as a continuous familiarity signal or as a continuous recollection-like signal. Either way, know judgments reflect weaker memory than remember judgments, and guess judgments reflect still weaker memory. Much evidence has accumulated in recent years bearing on the question of whether remember and know judgments reflect decisions based on recollection and familiarity, respectively, or whether they reflect different degrees of memory strength (as suggested by the signal-detection account). We turn now to a brief review of the evidence that seems to demand an interpretation in terms of memory strength, as illustrated in Figure 2.

### **Evidence That Remember/Know Judgments Reflect Memory Strength**

The idea that remember/know judgments might reflect different degrees of memory strength instead of different memory processes was recognized by dual-process theorists early on (e.g., Gardiner, 1988), but a variety of empirical dissociations were taken as evidence against the strength-based interpretation. For example, Gardiner and Java (1990) directly tested the idea that remember and know judgments were equivalent to sure and unsure judgments, respectively, as a strength-based interpretation would appear to assume (and as the signal-detection model would later appear to assume). In one experiment, they asked participants to make remember/know judgments following each old decision. In another, they asked for confidence ratings (sure/unsure). If remember judgments are equivalent to sure responses and know judgments are equivalent to unsure responses, then the results should be the same in both cases. Contrary to this prediction, participants made more know judgments than remember judgments to nonword target items, but they made more sure responses than unsure responses to nonword target items (findings that were recently

replicated by Geraci, McCabe, & Guillory, 2009). This dissociation has often been taken to mean that remember/know judgments are not equivalent to sure/unsure confidence ratings, and it was followed by other dissociations showing, for example, that manipulations thought to selectively affect recollection affected the remember hit rate while having little or no effect on the know hit rate (e.g., Gardiner, Gawlik, & Richardson-Klavehn, 1994).

As compelling as they might seem at first glance, dissociations such as these are easily reconciled with the signal-detection interpretation (Dunn, 2004; Hirshman & Master, 1997; Inoue & Bellezza, 1998; Wixted, 2009; Wixted & Stretch, 2004). For example, Wixted (2009) showed that the apparent dissociation for nonwords between remember/know judgments and confidence ratings reported by Gardiner and Java (1990) is readily accounted for by signal-detection theory once one also takes into consideration false alarm rates (instead of focusing exclusively on hit rates, as non-signal-detection theorists often do). No dissociation was observed for false alarm rates, which were higher for both know judgments and unsure judgments than for remember judgments and sure judgments. Moreover, the overall rate of false alarms was higher for sure judgments than for remember judgments (trends that were also observed by Geraci et al., 2009). This indicates that the sure criterion was placed in a more liberal position than the remember criterion. Once provisions are made for that difference, the signal-detection account predicts the observed dissociation for hits while also providing an interpretation of the absence of a crossover for false alarms. No dual-process theory makes simultaneous sense of the hit and false alarm rate patterns reported by Gardiner and Java (1990).

In a comprehensive review of the literature, Dunn (2004) showed that the vast majority of findings that were once declared to be problematic for the signal-detection point of view are in fact fully compatible with it. Furthermore, more recently, Dunn (2008) showed that it is not the case that the signal-detection account can explain any outcome. Instead, many experiments have used designs that could have yielded results that would be impossible for a signal-detection model to accommodate but could be easily accommodated by the dual-process interpretation. Even so, across many conditions from a variety of studies, outcomes incompatible with the signal-detection account were observed very rarely, and those few exceptions fell only slightly outside of the outcome space that can be accommodated by signal-detection theory. Thus, it is possible that even the few exceptions only reflect random error.

Wixted and Stretch (2004) reviewed a variety of other empirical findings that are compatible with the signal-detection interpretation and that are problematic for the dual-process perspective. Some of these issues concern the relationship between confidence ratings and remember/know judgments, an issue that will be of critical importance in the development of the new theory we present later. Figure 3 illustrates the signal-detection interpretation of that relationship. This figure corresponds to an old/new recognition task in which participants are asked to rate their confidence in old decisions on a 6-point scale (as in Figure 1) and to make a remember/know judgments as well. The findings discussed next are interpreted in relation to this figure.

**Confidence and accuracy.** Most obviously, the signal-detection account depicted in Figure 3 predicts that remember judgments will be made with higher confidence and higher accu-

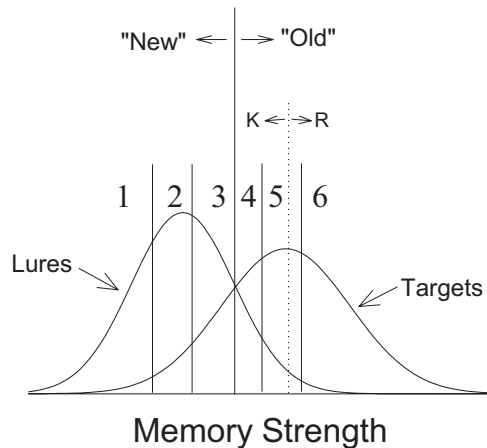


Figure 3. The signal-detection interpretation of a recognition memory task in which old/new confidence ratings are made on a 6-point scale and the participant is asked to make a remember/know judgment for items that receive a confidence rating of 4 or more ( $R = \text{remember}$  and  $K = \text{know}$ ). The dashed vertical line depicts the remember/know criterion.

racy than know judgments. As shown in the figure, remember judgments would be associated with confidence ratings of 5 or 6, whereas know judgments would be associated with confidence ratings of 4 or 5. In addition, accuracy for remember judgments is predicted to be higher than that for know judgments. This prediction arises because the proportion of the target distribution that falls to the right of the remember criterion is much greater than the proportion of the lure distribution that falls to the right of the remember criterion (which means that correct remember judgments will far exceed incorrect remember judgments). By contrast, the proportion of the target distribution that falls between the know criterion and the remember criterion is only somewhat greater for targets than for lures (which means that correct know judgments will only somewhat exceed incorrect know judgments). Almost invariably, remember judgments are made with higher confidence and higher accuracy than are know judgments (Wixted & Stretch, 2004).

For target items, higher confidence and accuracy for remember judgments relative to know judgments is demanded by the signal-detection view, but it can also be reconciled with the standard dual-process interpretation. That is, it seems reasonable to suppose that recollection-based memories (which are what remember hits are thought to reflect) would engender higher confidence and accuracy, on average, than familiarity-based memories (which are what know hits are thought to reflect). However, for lures, the two accounts make divergent predictions. With respect to confidence ratings, the signal-detection account makes the same predictions for false alarms as it does for hits. That is, as illustrated in Figure 3, remember false alarms should be made with higher confidence than know false alarms, a result that is almost invariably observed (e.g., Wixted & Stretch, 2004). By contrast, the dual-process interpretation of remember/know judgments is less clear about the nature of remember false alarms. Some accounts treat them as guesses (Yonelinas, 1994), whereas others tentatively treat them as false recollections (Diana, Reder, Arndt, & Park, 2006). If they are guesses, then they are guesses that are, for some reason, made with

high confidence. If they are assumed to reflect (categorical) false memories instead of guesses, then some of the criterion effects considered next become hard to explain.

**Criterion effects.** In the standard dual-process interpretation of remember/know judgments, a remember judgment is made when recollection occurs (i.e., recollection is an either/or categorical process). In the signal-detection interpretation, by contrast, a remember judgment is made when memory strength exceeds a high criterion. If the signal-detection interpretation is correct, then an experimental manipulation that induces a more liberal placement of that criterion (i.e., that induces a leftward shift of the remember criterion on the memory-strength axis while holding overall memory strength constant) should result in both a higher remember hit rate and a higher remember false alarm rate. Hirshman and Henzler (1998) used instructions to manipulate the placement of the remember criterion and showed that remember hit and false alarm rates covary in the expected manner. That is, remember hit and false alarm rates increase in response to instructions that are designed to induce a liberal criterion, whereas both decrease in response to instructions designed to induce a more conservative placement. Although this result is compatible with the signal-detection view, it could be argued that the remember/know instructions used in these studies were such that participants were induced to respond based on familiarity for both remember and know judgments.

A more compelling example of a remember criterion effect can be found by examining the remember hit and false alarm rates across participants instead of across biasing conditions. Often, the correlation is strongly and significantly positive, and this is even true of data that were advanced as evidence against the notion that a remember criterion exists. Dobbins, Khoe, Yonelinas, and Kroll (2000) showed that the overall hit rate is positively correlated with the overall false alarm rate across participants, a result they interpreted to reflect variations in the placement of the old/new decision criterion across participants. Wixted and Stretch (2004) further analyzed their data and found that, in addition, the remember hit rate correlated strongly with the remember false alarm rate across participants in all three of their experiments. This was also shown to be true of data previously reported by Stretch and Wixted (1998). This correlation suggests that participants who have a high remember hit rate also have a high remember false alarm rate, and the simplest explanation for that finding is the same one that applies to the old/new criterion. Specifically, participants vary in how liberal or conservative they are in where they place the remember criterion. If remember false alarms instead reflect false categorical recollection, it seems odd to suppose that participants who have a higher rate of true recollection are the very same participants who have a higher rate of false recollection. As such, if it is assumed that remember false alarms reflect false recollection, it must also be assumed that recollection is a continuous process and that participants use a criterion for deciding when to make a remember judgment, at which point one has essentially adopted a signal-detection interpretation of remember judgments much like the one that we later propose.

**Reaction time effects.** The signal-detection model does not offer a detailed account of reaction times, such as the shape of reaction time distributions and differences between correct and incorrect reaction times, as random walk and diffusion models do (Ratcliff & Starns, 2009). However, it has long provided a useful

conceptual framework for understanding basic effects on average reaction times (Norman & Wickelgren, 1969; Ratcliff & Murdock, 1976). In the signal-detection model, the old/new decision criterion is placed at the point of maximal subjective uncertainty. That is, from the participant's point of view, an item with memory strength that falls exactly at the old/new criterion is as likely to be a target as it is to be a lure. The reaction time rule derived from the signal-detection model holds that the speed of a recognition decision is inversely related to confidence (i.e., as confidence increases, reaction times decrease). The farther away from the old/new decision criterion memory strength falls in either direction (i.e., the farther from the point of subjective uncertainty), the faster the decision will be. Although a flatter function has been observed in highly practiced participants (Ratcliff & Starns, 2009), an inverse relation between confidence and reaction time is almost invariably observed in less practiced participants (whose performance presumably better reflects the routine behavioral tendencies they bring to the laboratory).

This simple strength-based conceptual scheme makes sense of remember/know reaction times. Because the memory strengths of items associated with know judgments are closer to the old/new criterion than the memory strengths of items associated with remember judgments (see Figures 2 and 3), know judgments should be made more slowly than remember judgments, a prediction that applies to both targets and lures. That is, remember hits should be faster than know hits, and remember false alarms should be faster than know false alarms. Moreover, because the memory strengths of lures that receive incorrect remember judgments are higher on the memory-strength scale than targets that receive correct know judgments, remember false alarms (which are incorrect decisions) should be made faster than know hits (which are correct decisions).

Much evidence has accumulated showing that the signal-detection reaction time predictions are all borne out. Wixted and Stretch (2004) reported that not only are remember hits faster than know hits but remember false alarms are also faster than know false alarms. In addition, remember *false alarms* are faster than know *hits*. This compelling pattern has been observed in numerous studies since that time (Dewhurst, Holmes, Brandt, & Dean, 2006; Duarte, Henson, & Graham, 2008; Wheeler & Buckner, 2004; Wiesmann & Ishaï, 2008). To take one example, Dewhurst et al. (2006) asked for old/new decisions followed by remember/know judgments and measured reaction times for the old/new decision. The reaction times associated with correct and incorrect remember judgments were 1,272 ms and 1,255 ms, respectively, whereas the reaction times associated with correct and incorrect know judgments were 2,015 ms and 1,982 ms, respectively.<sup>1</sup> Basically, the reaction time data accord with the idea that remember and know judgments reflect memories that differ in strength. In further support of this idea, Rotello and Zeng (2008) showed that reaction time differences for old/new decisions associated with remember and know judgments largely disappear once they are equated for confidence (a useful proxy for memory strength).

### Strength Is Not Enough

The various findings reviewed above (as well as other findings reviewed by Dunn, 2008) suggest that remember/know judgments reflect different degrees of memory strength, not recollection and

familiarity. Thus, a signal-detection model of some kind seems essential. However, the signal-detection view is not without its problems. In particular, how does the signal-detection model account for the butcher-on-the-bus phenomenon? If asked, would the individual who recognized the man on the bus in Mandler's classic anecdote—and recognized him with high confidence—give a remember judgment rather than a know judgment (as if one does not appreciate when strong recognition is based on familiarity)? Or is it instead the case that strong familiarity-based recognition memory is not something that ever actually occurs, in which case this predicament would not arise? The idea that it never occurs seems odd in light of subjective experience that suggests otherwise.

As we argue below, the key to reconciling signal-detection theory with dual-process theory and remember/know judgments comes from a detailed consideration of high-confidence familiarity-based recognition (i.e., the butcher-on-the-bus phenomenon). Indeed, a cogent criticism of the signal-detection interpretation of remember/know judgments leveled by Gardiner, Richardson-Klavehn, and Ramponi (1998)—one that signal-detection theorists have never addressed (until now)—is the following: “. . . there is no explanation of how a continuum of trace strength can give rise to qualitatively distinct kinds of mental experiences” (p. 285).

**Signal-detection theory and the butcher-on-the-bus phenomenon.** Whether memory strength in signal-detection theory is construed as an ahistorical familiarity variable (as it usually was in years gone by) or as an historical recollection-like variable (as seems more common today), interpretative complications arise in connection with the butcher-on-the-bus phenomenon. If the memory-strength variable in the signal-detection model is construed as familiarity, then the butcher-on-the-bus phenomenon is readily explained. According to this account, the phenomenon occurs when the familiarity of a test item exceeds a high criterion—high enough to warrant high confidence. Indeed, high-confidence, familiarity-based recognition is the very definition of the butcher-on-the-bus phenomenon. However, with regard to remember/know judgments, this account holds that, under those conditions, participants will not appreciate that their decision was based on familiarity, so, despite the absence of recollection, they will supply a remember judgment (claiming that the recollection of source detail was involved in their decision), not a know judgment. This seems like a potentially problematic idea.

If the memory-strength variable in the signal-detection model is instead construed as a recollection-like variable, then the very existence of the butcher-on-the-bus phenomenon becomes hard to explain. On this view, the continuous memory-strength variable is thought to contain increasing amounts of source information as memory strength increases. However, if strong, high-confidence recognition necessarily includes abundant source information, then the butcher-on-the-bus experience should not occur in the first place. The fact that it does seem to occur is therefore potentially problematic.

In contrast to the signal-detection accounts discussed above, the traditional dual-process interpretation holds that remember/know judgments are process-pure indicators of individual decisions

<sup>1</sup> These data were not reported in the article but were conveyed by Steven Dewhurst in a personal communication (August 20, 2008).

based on recollection or familiarity, not simply indirect indicators of memory strength. As such, the traditional account does not encounter these conceptual difficulties with respect to the butcher-on-the-bus phenomenon. Instead, it allows for the possibility of strong, familiarity-based memories, and it assumes that participants provide a know judgment under those circumstances (which implies that people appreciate when strong memory is accompanied by historical source information and when it is not). However, as indicated earlier, a problem for this view is that it seems basically untenable in light of overwhelming empirical evidence consistent with the strength-based interpretation illustrated in Figure 2.

Is there a solution to these various dilemmas—namely, a theory that is consistent with the empirical evidence (evidence that demands a strength-based interpretation of remember/know judgments), yet one that allows for the possibility of the butcher-on-the-bus phenomenon and that also acknowledges that people appreciate when they are experiencing high-confidence, familiarity-based memory? In what follows, we present a new dual-process signal-detection theory that, among other things, addresses these issues. The theory is entirely signal-detection based, but it nevertheless assumes that remember/know judgments can be used in such a way as to provide valid information about whether a decision concerning a high-strength item involved a high degree of recollection or not (contrary to the standard signal-detection view). All prior signal-detection-based criticisms of the remember/know procedure as it is typically used still apply, and the new account agrees with earlier signal-detection accounts suggesting that the standard remember/know procedure fails to distinguish between recollection and familiarity (despite its widespread use in that regard). However, the new model also points to a modified remember/know procedure that can be used productively to investigate recollection and familiarity. Of particular interest is the fact that the model provides an interpretation of know judgments that are occasionally made with high confidence (which is the experimental analog of the butcher-on-the-bus phenomenon). In our view, high-confidence know judgments warrant close attention, and we turn now to a consideration of how those judgments have been interpreted in the past.

**High-confidence know judgments.** Although most know judgments are associated with weaker memory than most remember judgments (as required by the signal-detection interpretation illustrated in Figures 2 and 3), not all know judgments reflect weak memory. That is, some know judgments are made with high confidence and, presumably, high accuracy. On the surface, high-confidence know judgments would seem to indicate the subjective experience that corresponds to the butcher-on-the-bus phenomenon.

The model illustrated in Figure 3 makes a clear prediction about the distribution of confidence ratings associated with remember/know judgments for an individual participant who is asked to make both confidence ratings and remember/know judgments for each item on a recognition test. The remember criterion in this example is placed between the criteria for making confidence ratings of 5 or 6. A participant whose decision criteria are arranged in this manner should provide confidence ratings of 4 or 5 for know judgments and provide confidence ratings of 5 or 6 for remember judgments. That is, some know judgments and some remember judgments would receive confidence ratings of 5, but no

know judgment would receive a confidence rating of 6, and no remember judgment would receive a confidence rating of 4. In fact, no matter where the remember criterion is placed with respect to the confidence criteria, remember and know judgments should share, at most, one confidence rating. If the remember criterion happened to be placed at precisely the same point as the one of the confidence criteria, then remember and know judgments would not share any confidence ratings across trials. For example, if the remember criterion were placed at precisely the same point as the criterion for making a confidence rating of 6, then all know judgments would be associated with confidence ratings of 4 or 5, and all remember judgments would be associated with confidence ratings of 6 (i.e., no overlap).

In practice, remember and know judgments from an individual participant often share more than one confidence rating (Wixted & Stretch, 2004). That is, a participant's remember judgments might consist mainly of confidence ratings of 6, but they may include some confidence ratings of 5, and a few confidence ratings of 4 may be evident as well. Most know judgments might consist mainly of confidence ratings of 4 and 5, but some confidence ratings of 6 may be observed as well. The standard signal-detection account does not anticipate this outcome. The high-confidence know judgments are of particular interest here because, on the surface, they signify the butcher-on-the-bus phenomenon.

Related to the confidence overlap problem is the fact that the slope of the z-ROC derived from confidence ratings may differ from the slope of the z-ROC derived from remember/know judgments (Rotello, Macmillan, & Reder, 2004). Specifically, the slope of the remember/know z-ROC is often closer to 1 than the slope of the confidence-based z-ROC. If confidence ratings and remember/know judgments exhibited the minimal overlap predicted by signal-detection theory, then the slopes of the two z-ROCs would be the same. But the overlap is often more extensive than the signal-detection account anticipates, so a model is needed that can explain the overlap and the related fact that the slope of the remember/know z-ROC is often close to 1.

To account for this ROC anomaly, Rotello et al. (2004) proposed a two-dimensional model instead (STREAK). This model assumes that an old/new decision for a particular test item is based on the sum of recollection and familiarity (with confidence being greater the greater the sum), whereas remember/know judgments are based on the difference between recollection and familiarity (with remember judgments being more likely the greater the difference). This model allows for extensive overlap between remember/know judgments and confidence ratings, and it naturally yields remember/know z-ROC slopes that are closer to 1 than the confidence-based z-ROC slope.

Wixted and Stretch (2004) took a different approach and explained both the confidence overlap and the differing z-ROC slopes on the basis of *criterion variability*. This approach retained the basic signal-detection model in that it assumes a unidimensional memory-strength axis. From a unidimensional perspective, the only way to explain the overlap between confidence ratings and remember/know judgments is to assume that either the confidence criteria or the remember criterion (or both) exhibit item-to-item variability with respect to each other. That is, for one test item, the criteria might be placed as shown in Figure 3. For another test item, either the confidence criteria or the remember criteria (or both) may shift such that the remember criterion falls above the

criterion for making a confidence rating of 6. This kind of instability in the placement of the criteria would allow remember/know judgments to share more than one confidence rating and could affect the slope of the z-ROC. In fact, if the remember criterion were more variable than the confidence criteria, then it would explain why the slope of the remember/know z-ROC is closer to 1 than the confidence-based z-ROC. That is, as criterion variability increases, the z-ROC will become closer to 1 (Benjamin, Diaz, & Wee, 2009). Thus, if the location of the remember criterion is more variable than the locations of the confidence criteria, the remember/know z-ROC slope should be closer to 1 than the confidence-based z-ROC (and extensive overlap between confidence ratings and remember/know judgments will be observed).

This simple criterion variability idea could explain why confidence ratings and remember/know judgments overlap as much as they do and why the slopes of confidence-based and remember/know-based z-ROCs differ. A number of empirical investigations have since shown that a unidimensional signal-detection model with criterion variability fits ROC data better than STREAK or alternative dual-process theories (e.g., Dougal & Rotello, 2007; Starns & Ratcliff, 2008). Rotello, Macmillan, Hicks, and Hautus (2006) found that data from about half of their participants were well fit by a fixed criterion model and half by a variable criterion model. The individual participants whose data were best fit by the variable criterion model tended to be the same participants with a remember/know z-ROC slope closer to 1 than the confidence-based z-ROC. This finding supports the idea that variability of the remember criterion is greater than the variability of the confidence criteria. Starns and Ratcliff (2008) specifically modeled variability for both the remember criterion and the confidence criteria and found that the estimated variability of the remember criterion significantly exceeded the estimated variability of the confidence criteria. All of these findings support the criterion-variability version of the unidimensional signal-detection interpretation of remember/know judgments.

For purposes of the present discussion, the critical point is that the criterion variability account does not regard the content of memory associated with a high-confidence remember judgment to be fundamentally different from the content of memory associated with a high-confidence know judgment. That is, the high-confidence memories can have the same content (i.e., a relatively high degree of source memory), but a very high remember criterion used on some test trials allows a know decision to be made. On other trials, a lower remember criterion is used, so equally strong memories with similar content receive a remember judgment instead.

Is the content of memory similar for high-confidence remember judgments and high-confidence know judgments? This question has never been addressed empirically. No existing signal-detection account anticipates that the amount of source information available is greater for remember judgments than for know judgments under those conditions. In what follows, a new signal-detection interpretation of remember/know judgments is presented, one that assumes a memory trace associated with a high-confidence know judgment is different from a memory trace associated with a high-confidence remember judgment in that the former is primarily based on familiarity (and corresponds to the butcher-on-the-bus phenomenon), whereas the latter involves a high degree of recollection. That is, this new model assumes that the butcher-on-the-bus phe-

nomenon is real and that participants supply a know judgment when it occurs. Three experiments are then presented to differentially test the criterion-variability account versus the account offered by the new model.

### A Continuous Dual-Process Signal-Detection Model

We refer to the new model as the continuous dual-process (CDP) model of recognition memory. The model is entirely signal-detection based, and it is unique in that it assumes that both recollection and familiarity are continuous signal-detection processes. The model was inspired by STREAK and has much in common with that model. For example, like STREAK, the CDP model assumes that (a) old/new recognition decisions (including confidence ratings) are based on an aggregated memory-strength variable composed of continuous recollection and familiarity signals (Wixted, 2007), and (b) remember and know judgments are based on different dimensions of memory strength instead of being made on the same dimension of memory strength (contrary to the model shown Figure 2). However, unlike STREAK, the CDP model does not envision a dimension upon which remember/know judgments are made that is independent of the aggregated memory-strength dimension. Instead, quite simply, the CDP model assumes that remember judgments are based on a recollection dimension and that know judgments are based on a familiarity dimension. Neither dimension is independent of the memory-strength variable; instead, the aggregated memory-strength variable upon which old/new decisions are based consists of the sum of the continuous recollection and familiarity signals upon which remember and know judgments are respectively made. Finally, like the standard dual-process interpretation of remember/know judgments, the CDP model assumes that individuals can report valid information about whether or not their decision was primarily based on recollection or familiarity using remember/know judgments. A critical difference, however, is that the remember/know judgments are not ever assumed to be process pure. Indeed, the model agrees with the main criticism leveled by signal-detection theorists in recent years: Remember/know judgments, as ordinarily used, reflect different degrees of memory strength and do not effectively disentangle recollection and familiarity.

### A Verbal Description and Visual Illustration of the CDP Model

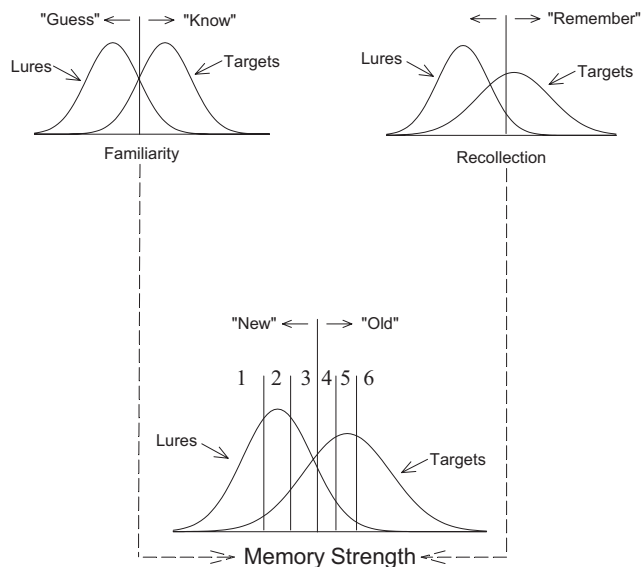
In the CDP model, each recognition memory process has its own "noise" distribution (for lures) and its own "signal" distribution (for targets). This idea is what separates the new model from all prior dual-process models, and it is what brings it into harmony with one particular use of remember/know judgments for distinguishing recollection-based and familiarity-based memories. An old/new recognition decision is assumed to be based on the sum of the recollection and familiarity signals for a particular test item. That is, the memory-strength signal for a test item consists of recollection and familiarity combined, and this is true of both targets and lures. The confidence criteria are placed along the aggregated memory-strength axis (as in Figure 1). However, in the CDP model, the remember and know criteria are not placed on the same memory-strength axis (i.e., the situation is not as depicted in Figures 2 and 3). Instead, the remember criterion is placed on



the recollection axis, and the know criterion is placed on the familiarity axis. The CDP model is illustrated in Figure 4.

For an item with an aggregated memory strength that exceeds the old/new decision criterion (and therefore receives a confidence rating of 4, 5 or 6), the participant is assumed to make a remember/know judgment by first interrogating the recollection dimension by asking, in effect, "How much source information do I recollect about this item?" In this account, recollection is a continuous signal, as much evidence suggests is the case (e.g., Johnson, McDuff, Rugg, & Norman, 2009; Kurilla & Westerman, 2010; Mickes, Johnson, & Wixted, 2010; Mickes, Wais, & Wixted, 2009; Palmer, Brewer, McKinnon, & Weber, 2010; Ratcliff & Starns, 2009; Slotnick, 2010; Starns, Hicks, Brown, & Martin, 2008; but see Parks & Yonelinas, 2009, for evidence that recollection is a threshold process). A remember judgment is made if the participant is satisfied that *enough* recollection has occurred. That is, a remember judgment is made if the strength of the recollection signal exceeds the remember criterion. Different participants would be expected to use different settings for the remember criterion, so the remember hit rate should be correlated with the remember false alarm rate across participants, as is typically the case (Wixted & Stretch, 2004). Note that this differs from the more common process-pure interpretation, which is that remember judgments reflect the occurrence of recollection (rather than enough recollection) and that lures do not generate a recollection signal.

In the CDP model, if the degree of recollection fails to exceed the remember criterion, then the participant makes a know judg-



**Figure 4.** An illustration of the continuous dual-process signal-detection (DPSD) model. For old/new decisions, the recollection and familiarity signals are assumed to be summed, so the model reduces to the standard unequal-variance signal-detection account illustrated in Figure 1. However, the participant is asked to make a remember/know/guess judgment, memory is then queried for recollection, and the participant makes a remember judgment if the recollection signal exceeds a decision criterion. If recollection fails to exceed that criterion, memory is queried for familiarity, and the participant makes a know judgment if familiarity exceeds a criterion (otherwise, a guess judgment is made).

ment (if the only options are remember and know) or interrogates the familiarity dimension (if the options are remember, know or guess). The familiarity signal is also considered to be continuous, in agreement with all dual-process models. If the degree of familiarity exceeds the know criterion (i.e., if enough familiarity occurs), a know judgment is made; otherwise, a guess judgment is made.

Note that a know judgment in this model implies that the amount of familiarity associated with the test item exceeds a criterion on the familiarity dimension, but it does not imply the absence of recollection. Instead, the amount of recollection is such that it simply failed to exceed the criterion on the recollection dimension. Recent evidence supports the view that know judgments entail measurable degrees of recollection, not the absence of recollection (Eldridge, Engel, Zeineh, Bookheimer, & Knowlton, 2005; Hicks, Marsh, & Ritschel, 2002; Johnson et al., 2009; Wais, Mickes, & Wixted, 2008). For example, Wais et al. (2008) found that source recollection was greater for remember judgments than for know judgments, but they also found that recollection was above chance for items judged to be known (even when participants were asked to make a source-recollection decision before making a remember/know judgment). In addition, Johnson et al. (2009) found that the cortical reinstatement of encoding activity at retrieval (thought to be the neural signature of recollection) was higher for remember judgments than for know judgments, but reinstatement was evident for both.

Although the CDP model assumes that participants can separately query the recollection and familiarity dimensions, it also assumes that they do not necessarily do so when asked to decide whether an item is old or new. Instead, under those conditions, the familiarity and recollection values are summed to create a memory-strength variable, and the decision is made on the basis of that aggregated variable with respect to the confidence criteria. If the recollection and familiarity signals are not completely redundant, and if both signals are reliable indicators of prior occurrence, it is often the case that the aggregated variable will be more diagnostic of prior occurrence than either variable considered alone. That being the case, it would be reasonable to sum them.

If the recollection and familiarity values are summed, and if the target distribution for at least one process has greater variance than its corresponding lure distribution, then, for old/new decisions, the model reduces to the traditional unequal-variance signal-detection model. This is an important feature of the model in light of a large body of evidence supporting the signal-detection view of confidence ratings over competing models of a similar level of complexity (e.g., Jang, Wixted, & Huber, 2009; Rotello et al., 2006; Slotnick & Dodson, 2005; Starns & Ratcliff, 2008). By making the assumption that recollection and familiarity are summed, the CDP model is brought into harmony with that body of evidence. This assumption seems particularly defensible when participants are asked to make old/new judgments and confidence ratings in the absence of remember/know judgments (e.g., in a typical experiment on confidence-based ROCs). However, the question of whether or not the two processes are summed is an empirical one that has not yet been specifically addressed. Indeed, it is conceivable that asking participants to make remember/know judgments for every test item can induce a strategy of separately checking for enough recollection and then, if necessary, checking for enough familiarity even before making the old/new confidence rating

(instead of summing recollection and familiarity before making that rating). Nevertheless, for the sake of simplicity, we generally make the simplifying assumption that recollection and familiarity are summed to make old/new confidence ratings even when the remember/know procedure is used.

### Relationship to Other Models

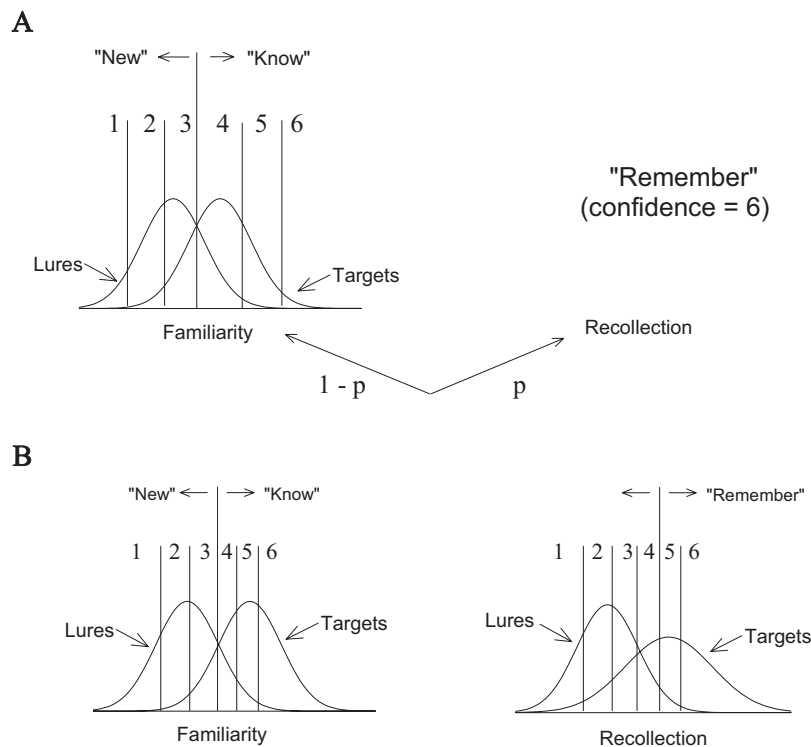
As discussed above, the CDP model is similar to STREAK, but, in some ways, it is similar to the dual-process signal-detection (DPSD) model proposed by Yonelinas (1994) as well. To emphasize the key differences between the two models, the DPSD model is illustrated in Figure 5A, and another version of the CDP model is shown in Figure 5B. This version would apply if asking for remember/know judgments for every item induces participants to query the recollection dimension and then, if necessary, the familiarity dimension even when making an old/new confidence rating. This version of the model is the most similar to the DPSD model, which allows for their key differences to be highlighted.

Unlike the CDP model, the DPSD model assumes that recollection is a categorical process (sometimes referred to as a “threshold” process), which means that for a particular test item, recol-

lection either occurs (in which case confidence is high, and a remember judgment is made) or does not occur (in which case the decision is based on familiarity). For lures, the probability of recollection is zero, and for targets, the probability of recollection is greater than zero (and is represented by  $p$  in the figure). In the absence of recollection, the level of confidence for a particular test item is based on its degree of familiarity, and a know judgment is made if familiarity is sufficiently high (otherwise the decision is new). Note that, in this model, any level of confidence less than 6 reflects a decision that is based purely on familiarity.

It has been suggested that the DPSD model allows for a continuous recollection signal in the sense that one can recollect a little or a lot about a prior episode (Parks & Yonelinas, 2007). However, this assumption does not affect how the DPSD model views the relationship between recollection and confidence, because the model assumes that any degree of recollection always yields high confidence. The model further holds that lures are not associated with any degree of recollection.

Both the DPSD model and the CDP model assume that familiarity is a continuous signal-detection process in which the mean of the target distribution is greater than the mean of the lure distri-



*Figure 5.* A. An illustration of the dual-process signal-detection (DPSD) model when a 6-point confidence scale is used. Whether making old/new confidence ratings or remember/know judgments, the model assumes that the recognition decision for a test item is based on recollection if recollection occurs (in which case confidence is rated 6, and a remember judgment is made) or is based on familiarity if recollection fails (in which case the confidence rating depends on the level of familiarity, and a know judgment is made if familiarity is high enough to make a confidence rating of at least 4). The probability that recollection occurs for a test item is  $p$ . B. An illustration of a version of the continuous dual-process (CDP) model that assumes that recollection and familiarity are separately queried even to make old/new confidence ratings. In this version, recollection and familiarity are not summed, and separate old/new confidence criteria are placed on the recollection and familiarity axes.

bution, but their variances are equal. The equal-variance assumption is a defining feature of the DPSD model, and, for sake of simplicity, we adopt the same assumption for the CDP model. However, the equal-variance assumption is a much more critical assumption for the DPSD model than it is for the CDP model. For example, according to the DPSD model, a symmetrical confidence-based ROC necessarily implies familiarity-based responding. According to the CDP model, it does not necessarily imply familiarity-based responding. Although we assume that the distributions for recollection have unequal variance, the CDP model allows for the possibility that they can have equal variance (such as when memory is weak). In that case, the confidence-based ROC would be symmetrical even though decisions are based on both recollection and familiarity.

The main difference between the DPSD model and the CDP model is that the DPSD model views recollection as a categorical process, whereas the CDP model views recollection as a continuous signal-detection process (one that is associated with varying degrees of confidence and varying degrees of accuracy). If recollection is a continuous signal-detection process, then a recollection signal is associated with both targets and lures (not just with targets). This is why, in Figure 4 and in Figure 5B, the recollection model has a lure distribution as well as a target distribution.

A lure distribution for recollection may seem odd to those accustomed to thinking of recollection as something that can occur only for target items. Intuition would suggest that a recollection signal should not occur for items that did not appear on the list, because it does not seem possible to recollect an experience that did not occur (with the possible exception of a few false memories that may be created during list presentation). In a signal-detection model, however, recollection is a signal that occurs in response to a query of memory, and it occurs for every test item. Because retrieval from memory is a noisy process, even lures are associated with a recollection signal despite the fact that no source information was encoded. This idea is analogous to the more familiar concept of perceptual noise. For example, someone who is waiting for a call on a cell phone set to vibrate will sometimes mistakenly sense vibration and, therefore, answer a call that has not been made. An analogous process is assumed to occur when someone attempts to decide if a particular test item (which may be a lure) is associated with recollection. That is, if memory is queried for evidence of recollection (and if one is alert to the slightest trace of a recollection signal), then mnemonic noise may indicate that recollection has occurred even if no information about the test item was encoded at the time of study. This is why a recollection signal is associated not just with targets that were encoded during study but also with lures (and with targets that may not have been encoded at study).

These considerations separate the CDP model even from other models that have allowed for the possibility of a continuous recollection signal. The variable recollection dual-process (VRDP) model, for example, is much like the DPSD model in that it assumes that recollection occurs for targets but not for lures (Onyper, Zhang, & Howard, 2010; Sherman, Atri, Hasselmo, Stern, & Howard, 2003). It differs mainly in that it assumes that recollection is a continuous process that can be associated with varying degrees of confidence (not just with the highest level of confidence). In that sense, it is like the CDP model. However, because it assumes that no recollection signal is returned by lures,

the VRDP model views recollection as a property of the encoded trace. Thus, it does not view recollection as a signal-detection process, and, as such, it does not contain provisions for a lure distribution associated with recollection (e.g., see Figure 2D in Onyper et al., 2010). By contrast, the CDP model views recollection as a signal that occurs in response to a retrieval cue, and it assumes that a recollection signal of some magnitude (however small) is returned whenever memory is queried for evidence of recollection.

### A Simulation-Based Illustration of the CDP Signal-Detection Model

In this section, the performance of the CDP model is illustrated by means of a simple Monte Carlo simulation, and we begin by introducing the mathematical notation that is used to explain how the simulation was performed. The familiarity process involves a lure distribution with mean and standard deviation that are henceforth represented by  $\mu_{F0}$  and  $\sigma_{F0}$  (the subscript F denotes familiarity, and the subscript 0 denotes the “noise” or lure distribution), respectively, and a target distribution with mean and standard deviation that are henceforth represented by  $\mu_{F1}$  and  $\sigma_{F1}$ , respectively (the subscript 1 denotes the “signal plus noise” or target distribution). The distance between the two distributions (i.e., the ability to discriminate targets from lures on the basis of familiarity alone) would equal  $d'_{F}$  for the case in which  $\sigma_{F0} = \sigma_{F1}$  (and would instead be represented by  $d_{a-F}$  if the variances are unequal). Similarly, the recollection process involves a lure distribution with mean and standard deviation represented by  $\mu_{R0}$  and  $\sigma_{R0}$ , respectively, and a target distribution with mean and standard deviation represented by  $\mu_{R1}$  and  $\sigma_{R1}$ , respectively. The distance between the two distributions (i.e., the ability to discriminate targets from lures on the basis of recollection alone) would equal  $d'_{R}$  for the case in which  $\sigma_{R0} = \sigma_{R1}$  (and would be represented by  $d_{a-R}$  if the variances are unequal).

For this simulation, the means of the familiarity and recollection lure distributions ( $\mu_{F0}$  and  $\mu_{R0}$ , respectively) were both set to 0, and their corresponding standard deviations ( $\sigma_{F0}$  and  $\sigma_{R0}$ , respectively) were both set to 1. The means of the familiarity and recollection target distributions ( $\mu_{F1}$  and  $\mu_{R1}$ , respectively) were set to 0.80 and 1.0, respectively, and their corresponding standard deviations ( $\sigma_{F1}$  and  $\sigma_{R1}$ , respectively) were set to 1.0 and 1.4, respectively. Thus, an equal-variance model was assumed to underlie the familiarity process. This matches an assumption made by the DPSD model proposed by Yonelinas (1994), but it is not a necessary assumption of the CDP model. On the recollection dimension, the remember criterion was placed 1.8 standard deviations above the mean of the recollection lure distribution ( $\mu_{R0}$ ). On the familiarity dimension, the know criterion was placed 1.5 standard deviations above the mean of the familiarity lure distribution ( $\mu_{F0}$ ).

For old/new decisions, values drawn from the recollection and familiarity dimensions were summed. Thus, the mean of the target and lure distributions upon which old new decisions are based are given by

$$\mu_{\text{Target}} = \mu_{F1} + \mu_{R1},$$

$$\mu_{\text{Lure}} = \mu_{F0} + \mu_{R0},$$

and the corresponding variances are given by

$$\sigma_{\text{Target}}^2 = \sigma_{F1}^2 + \sigma_{R1}^2 + 2\rho_1 \cdot \sigma_{F1} \cdot \sigma_{R1},$$

$$\sigma_{\text{Lure}}^2 = \sigma_{F0}^2 + \sigma_{R0}^2 + 2\rho_0 \cdot \sigma_{F0} \cdot \sigma_{R0},$$

where  $\rho_1$  is the correlation between the recollection and familiarity values of the targets, and  $\rho_0$  is the correlation between the recollection and familiarity values of the lures. The ability to discriminate between targets and lures is given by  $d_a$ , which is equal to

$$d_a = \frac{\mu_{\text{Target}} - \mu_{\text{Lure}}}{\sqrt{(\sigma_{\text{Target}}^2 + \sigma_{\text{Lure}}^2)/2}}$$

If we assume that  $\rho_1 = \rho_0 = 0$  (i.e., the recollection and familiarity signals are uncorrelated for both targets and lures), and keeping in mind that  $\mu_{F0}$  and  $\mu_{R0}$  were both arbitrarily set to 0 and that  $\sigma_{F0}^2$  and  $\sigma_{R0}^2$  were arbitrarily set equal to 1, this equation reduces to

$$d_a = \frac{\mu_{F1} + \mu_{R1}}{\sqrt{(\sigma_{F1}^2 + \sigma_{R1}^2 + 2)/2}}$$

Holding all else equal, this equation states that the ability to discriminate targets from lures can be increased by increasing either the mean of the familiarity target distribution ( $\mu_{F1}$ ) or the mean of the recollection target distribution ( $\mu_{R1}$ ).

Summing recollection and familiarity when making old/new decisions effectively assigns equal weight to both processes. It seems reasonable to suppose that participants could assign different weights if they are instructed to ignore one process (e.g., Hockley & Cristi, 1996; Migo, Montaldi, Norman, Quamme, & Mayes, 2009) or if they learn through trial and error that a particular experimental task can be solved more efficiently on the basis of one process or the other. For the sake of simplicity, however, we assume equal weighting of recollection and familiarity in the basic version of the CDP model.

Note that  $d_{a-R}$  (discriminability in the recollection channel) is equal to  $\mu_{R1}/\sqrt{(\sigma_{R1}^2 + 1)/2}$  and that  $d_{a-F}$  (discriminability in the familiarity channel) is equal to  $\mu_{F1}/\sqrt{(\sigma_{F1}^2 + 1)/2}$ . For the values used above,  $d_{a-R} = 0.82$  and  $d_{a-F} = .80$ . However,  $d_a$  (discriminability based on recollection and familiarity combined) is equal to 1.14. Thus, for these settings, relying on recollection and familiarity combined would lead to better discriminability than relying exclusively on one process or the other. The higher the correlation between the recollection and familiarity signals (i.e., the more redundant those signals are), the less the advantage of combining them. For example, if  $\rho_1$  and  $\rho_0$  both equal 0.5, then  $d_a$  drops to 0.94. We make the assumption that, in everyday life, the combined signal is more diagnostic than either signal alone, in which case participants would learn to combine them as a matter of course. However, the model does not mathematically require that the signals be combined (as in Figure 5B). In addition, because it assumes that participants are capable of separately interrogating the recollection and familiarity dimensions, the CDP model implies that participants are capable of deemphasizing one process in favor of the other. Still, by assuming that the recollection and familiarity signals are combined, the model parsimoniously yields the longstanding unequal-variance signal-detection model of old/new recognition memory.

If a 6-point confidence scale is used for old/new decisions (as is commonly done), five confidence criteria would need to be placed

along the aggregated memory-strength axis (as in Figure 1). For this illustrative version of the model, those five criteria were placed in relation to  $\mu_{\text{Lure}}$  in the following locations (with the units equal to  $\sigma_{\text{Lure}}$ ): 2.0, 1.5, 1.0, 0.50, and 0.

Next, simulated data were generated by drawing 1,000 values from the lure distribution and 1,000 values from the target distribution. Thus, for example, for a lure, a familiarity value was randomly drawn from the lure distribution in the familiarity channel ( $f_{oi}$ ), and a recollection value was randomly drawn from the lure distribution in the recollection channel ( $r_{oi}$ ). These values were summed to generate an aggregated memory-strength value for that lure,  $m_{oi}$  (i.e.,  $m_{oi} = f_{oi} + r_{oi}$ ). The old/new confidence rating for this simulated test item is equal to the highest confidence criterion exceeded by  $m_{oi}$ . If  $m_{oi}$  did not exceed the old/new criterion (i.e., if it did not exceed the confidence criterion for a confidence rating of 4), a “new” decision was made, the confidence rating (1, 2, or 3) was recorded, and the next test item was considered. If  $m_{oi}$  instead exceeded the old/new criterion (i.e., if it exceeded the confidence criterion for a rating of 4 or more), then an “old” decision was made, the confidence rating (4, 5, or 6) was recorded, and a remember, know, or guess judgment was made. To make that judgment,  $r_{oi}$  (the recollection value) was first checked against the remember criterion on the recollection dimension. If its value exceeded the remember criterion (i.e., if this item generated a strong enough sense of recollection), then a remember judgment was recorded. Because the item in this example is a lure, this would be a remember false alarm. If its value did not exceed the remember criterion,  $f_{oi}$  was then checked against the know criterion on the familiarity dimension. If its value exceeded the know criterion (i.e., if this item generated a strong enough sense of familiarity), a know judgment was recorded. If its value did not exceed the know criterion, a guess judgment was made. Once this judgment was made, the next test item was considered (and so on for 1,000 targets and 1,000 lures).

Although neither the confidence criteria nor the remember and know criteria are assumed to be variable, the model nevertheless generates data that create the appearance of criterion variability from the point of view of the unidimensional model. That is, remember and know judgments generally share more than one confidence rating. Moreover, the model often yields a remember/know z-ROC slope that is closer to 1 than the confidence-based z-ROC. Table 2 shows the distribution of confidence ratings as-

Table 2  
*Distribution of Simulated Responses for Targets and Lures as a Function of Confidence Rating and Remember/Know/Guess (R, K, and G, Respectively) Judgment*

Confidence	Target				Lure			
	R	K	G	New	R	K	G	New
1	—	—	—	146	—	—	—	500
2	—	—	—	115	—	—	—	192
3	—	—	—	149	—	—	—	149
4	25	31	107	—	9	17	67	—
5	47	42	60	—	9	15	19	—
6	198	68	10	—	10	11	2	—

Note. Dashes indicate no response.

sociated with remember and know judgments for this simulation, and it is clear that they share more than one confidence rating. This is the kind of result that has, in the past, been explained in terms of criterion variability (Wixted & Stretch, 2004).

Figure 6A shows the average confidence associated with the simulated remember, know, and guess judgments, and Figure 6B shows the average old/new accuracy associated with the simulated remember, know, and guess judgments. Both plots exhibited the graded strength pattern that has, in the past, been explained in terms of the standard signal-detection interpretation of remember/know judgments shown in Figures 2 and 3. These simulation results show that the same strength-related effects are consistent with the CDP model shown in Figure 4. In addition, this simulated run produced a confidence-based ROC slope of 0.82 and a remember/know/guess ROC slope of 0.92, which is the kind of result that has in the past been explained in terms of the standard signal-detection model with added criterion variability. The relevant z-ROC plots are shown in Figures 6C and 6D.

In the simulation described above, the recollection and familiarity signals were assumed to be uncorrelated. That is,  $\rho_1$  and  $\rho_0$  (the correlation between the recollection and familiarity signals for targets and lures, respectively) were both set to 0. As that correlation increases, the predicted overlap between confidence ratings and remember/know/guess judgments decreases. In fact, when the correlation for both processes is 1, the CDP model reduces to the standard signal detection remember/know model illustrated in Figure 3. Under those conditions, the slope of the z-ROC based on confidence ratings would match the slope of the z-ROC based on remember/know judgments. Another way that this would occur is if a participant made all recognition decisions using only one process. For example, if a participant responded solely on the basis of the recollection signal (ignoring familiarity), then, for that participant, the CDP model would be equivalent to the standard unidimensional signal-detection model. The ability of the CDP model to reduce to the standard signal-detection model seems important in light of evidence showing that one third to one half of participants yield remember/know and confidence-rating data that

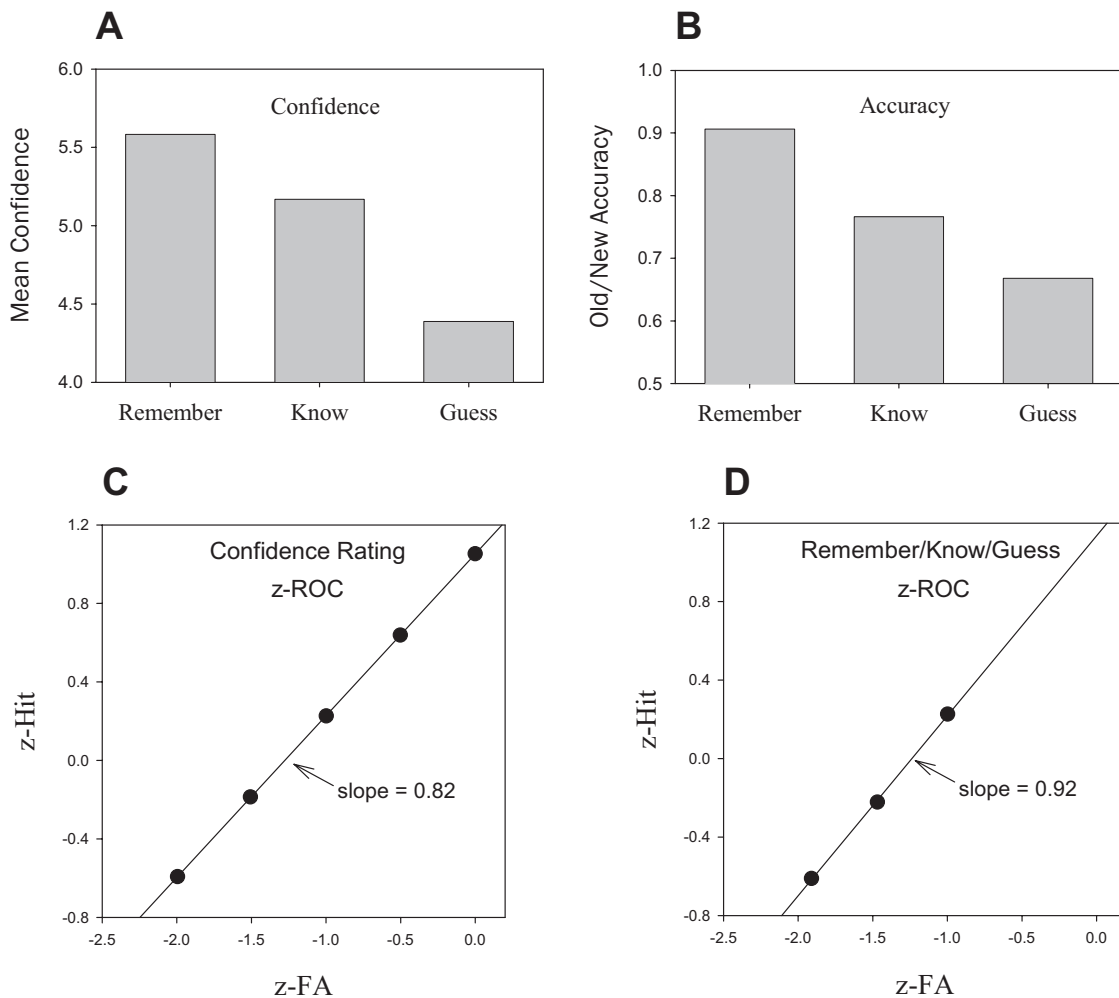


Figure 6. Mean confidence (A) and old/new accuracy (B) scores based on a representative Monte Carlo simulation of the continuous dual-process model. The confidence-based z-ROC (C) has a slope of 0.82, but the remember/know/guess z-ROC (D) has a slope of 0.92.

are consistent with that model (Rotello et al., 2006; Wixted & Stretch, 2004). The remaining participants exhibit the kind of overlap discussed above, which can be accounted for either in terms of criterion variability or in terms of the CDP model.

Note that the higher the correlation between recollection and familiarity signals, the less likely it is that a high-confidence know judgment will occur. If the correlation between the two signals is 1.0, then strong familiarity would necessarily imply strong recollection as well, so there would be no occurrences of strong memory involving a high degree of familiarity and a low degree of recollection (which, in turn, means that there would be no high-confidence know judgments). Thus, the butcher-on-the-bus phenomenon would never occur, as it apparently does not for those participants whose data are adequately characterized by the Donaldson (1996) signal-detection model illustrated in Figure 3.

### Fitting the CDP Model to Empirical Data

Next, we fit the CDP model and the unidimensional signal-detection model (with criterion variability) to experimental data reported by Rotello, Macmillan, Reeder, and Wong (2005). In that study, participants studied a list of words and then, on a subsequent recognition test, were asked to supply both confidence ratings (on a 6-point scale) and remember/know/new judgments. A remember/know/new judgment was requested for every test item that received a confidence rating of 2 or greater on the 6-point confidence scale. Participants performed this task in either the neutral condition or the conservative condition. The neutral condition used standard remember/know instructions, whereas the conservative condition used instructions designed to induce a more conservative placement of the remember criterion (resulting in fewer remember judgments to both targets and lures). The conservative placement of the remember criterion was induced by informing participants that they could be asked to justify any remember judgment they made. We fit the criterion-variability model and the CDP model to the group data from both conditions. For comparative purposes, we also fit STREAK. Dougal and Rotello (2007) already showed that STREAK has difficulty fitting data such as these, but we included STREAK in these fits because the CDP model shares much in common with STREAK, and the inclusion of that model underscores the point that it is not trivial for a model to adequately fit data such as these. We did not fit the DPSD model because that model is qualitatively contradicted by data like these, which show that remember judgments can occur over a wide range of confidence (not just for the highest level of confidence, as the DPSD model requires).

The group data reported by Rotello et al. (2005) exhibit considerable overlap in confidence ratings and remember/know judgments (such that remember/know judgments share more than one confidence category). Although such overlap can occur in group data even if every single individual participant shows no overlap, in practice, many participants exhibit considerable overlap. For those who do not, both the criterion-variability model and the CDP model can reduce to the standard signal-detection model shown in Figures 2 and 3 (a model that predicts that remember/know judgments share, at most, one confidence rating). The criterion-variability model reduces to the standard signal-detection model when criterion variability drops to zero. The CDP model reduces to the standard model when the correlation between the recollec-

tion and familiarity signals is 1.0 (or when participants use only one process to make their recognition decisions).

The purpose of this model-fitting exercise was not to establish which model provides a better fit of the data. A convincing test of that nature would require fits at the level of the individual participant with adjustments made for differences in model flexibility across multiple sets of data (e.g., Cohen, Rotello, & Macmillan, 2008; Jang, Wixted, & Huber, 2009). Our purpose was simply to show that the CDP model can provide a reasonably accurate quantitative description of group data that are representative of individuals who show overlap in confidence ratings and remember/know judgments (as many but not all participants do). If the CDP model were unable to do that, its viability would be open to question at the outset.

All three models were fit to the neutral and conservative data reported by Rotello et al. (2005) in their Table A1. The models were fit using maximum likelihood estimation, and the equations used to perform these fits for the criterion variability and CDP models are presented in the Appendix. Each model involved nine parameters. For all three models, six of the nine parameters consisted of the locations of decision criteria (five confidence criteria and one remember criterion). For the criterion-variability model, the other three parameters consisted of the mean of the target distribution, the standard deviation of the target distribution (relative to the lure distribution), and the standard deviation of the remember criterion (because the remember criterion was not fixed). For the CDP model, the other three parameters consisted of the mean and standard deviation of the target distribution for recollection and the mean of the target distribution for familiarity (its standard deviation was fixed at 1.0, as in the simulation described above). For STREAK, the other three parameters were  $d'_x$  (global memory strength, or familiarity),  $d'_y$  (specific memory strength, or recollection), and  $s$  (the standard deviation of the lure distribution relative to the target distribution). More parameters could be added to each model, but these nine-parameter versions serve to illustrate the basic performance of each model in relation to empirical data.

Table 3 shows the results of these fits. The table shows the estimated parameter values for each model and also shows the corresponding chi-square goodness-of-fit statistic. For both conditions, STREAK provides a poor fit, which others have reported as well (e.g., Dougal & Rotello, 2007; Starns & Ratcliff, 2008). The reason why STREAK has difficulty with data such as these is that it does not expect any particular relationship between average confidence ratings and remember/know judgments (i.e., the model does not predict that remember judgments will be made with higher confidence than know judgments). Thus, the strong relationship between confidence ratings and remember/know judgments evident in the data (i.e., remember judgments are invariably associated with higher average confidence than are know judgments) leads to a poor fit. The criterion variability and CDP models both provide much better fits that are comparable to each other.

All of the models interpret the result of the experimental manipulation in a similar way. That is, in all three models, the effect of the experimental manipulation was to induce a more conservative placement of the remember criterion. However, other parameters also appear to be affected by the experimental manipulation as well, although to lesser degrees. For example, all three models

Table 3  
Best-Fitting Parameter Values and Chi-Square Goodness-of-Fit Statistics for Three Models Fit to the Group ROC Data Reported by Rotello et al. (2005)

Parameter	STREAK		Parameter	Criterion variability model		Parameter	CDP model	
	Neutral	Conservative		Neutral	Conservative		Neutral	Conservative
$c_6$	1.28	1.19	$c_6$	1.68	1.77	$c_6$	2.32	2.56
$c_5$	0.92	0.94	$c_5$	1.20	1.39	$c_5$	1.67	2.00
$c_4$	0.63	0.74	$c_4$	0.82	1.08	$c_4$	1.15	1.54
$c_3$	0.37	0.48	$c_3$	0.48	0.70	$c_3$	0.69	0.99
$c_2$	0.04	0.18	$c_2$	0.05	0.27	$c_2$	0.08	0.36
$r$	0.21	-0.12	$\mu_r$	1.28	1.92	$r$	0.92	1.58
$d_x$	0.50	0.43	$\sigma_r$	0.96	0.74	$\mu_{R1}$	0.81	0.70
$d_y$	0.90	0.89	$d$	1.15	1.16	$\mu_{F1}$	0.78	0.96
$s$	0.76	0.68	$\sigma_{\text{Target}}$	1.33	1.51	$\sigma_{R1}$	1.48	1.98
$\chi^2$	493.67	595.64	$\chi^2$	10.98	24.39	$\chi^2$	12.15	21.80

Note. ROC = receiver operating characteristic; CDP = continuous dual-process.

suggest that the confidence criteria were more conservatively placed in the conservative condition compared to the neutral condition. The CDP model suggests that this effect was somewhat more pronounced than the other two models.

As indicated earlier, both models can account for the data of individual participants who yield data consistent with the standard signal-detection model depicted in Figure 3. It is conceivable that, for the remaining participants, it will be possible to distinguish the two models based on goodness-of-fit statistics, but we have so far found that data generated by one model can be readily fit by the other. A potentially better approach to distinguishing between the two accounts would be to test their predictions about memory performance. After all, making sense of empirical data (in addition to fitting such data with the highest degree of accuracy) is an important job of each model. We turn now to three experiments designed to test novel predictions of the CDP model.

### Experimental Tests of CDP Signal-Detection Model

The most direct way to investigate the validity of the CDP model is to determine what high-confidence know judgments represent. The frequency with which these judgments occur varies considerably across studies, but the fact that they do occur is not in question; only their meaning is. In a criterion variability account, the content of memory for old/new decisions made with the same confidence and accuracy but with different remember/know judgments would not necessarily be expected to differ. Instead, the differing judgments would be attributable to the decision process. Thus, for example, two items associated with an old/new confidence rating of 6 (i.e., old-6), one of which receives a know judgment and the other of which receives a remember judgment, could have similar source-memory content. Even so, one old-6 item would receive a know judgment because the remember criterion happened to be placed at an especially conservative position on that trial. Another old-6 item would instead receive a remember judgment because, on that trial, the remember criterion happened to be placed at a lower point on the memory strength axis. Across all trials, if old/new accuracy for high-confidence know judgments equaled old/new accuracy for high-confidence remember judgments (i.e., if memory strength were equated in terms of confi-

dence and old/new accuracy), then source accuracy might be equated as well. Such an outcome would be easily reconciled with the criterion variability account. Indeed, in a conceptually similar approach, Rotello and Zeng (2008) found that the usual reaction time differences between remember and know judgments all but disappeared after equating for confidence, and they appealed to the criterion variability account to explain their findings. If the same result occurs for source recollection, it would further weigh against the notion that remember/know judgments indicate anything about recollection and familiarity once memory strength is equated (and it would render the CDP model unnecessary).

The CDP model predicts that items with the same old/new strength sometimes receive different remember/know judgments because a remember judgment identifies an old decision that involves a considerable degree of recollection, whereas a know judgment identifies an old decision that involves a lesser degree of recollection (although not the absence of recollection). Thus, know judgments for old decisions made with high confidence and high accuracy (a subset of know judgments) reflect decisions based largely on a strong sense of familiarity. These decisions correspond closely to Mandler's butcher-on-the-bus phenomenon.

The predictions of the criterion variability and CDP models not only differ from each other, they also differ from the prediction made by the standard process-pure interpretation of remember/know judgments. In the process-pure account, a high-confidence remember judgment should be associated with considerable source recollection, but a high-confidence know judgment, being purely familiarity based, should be devoid of source recollection (i.e., source memory should be at chance). Prior research does not support the latter prediction (e.g., Wais et al., 2008). One could argue that source-recollection accuracy associated with know judgments reflects "unitized familiarity" instead of continuous recollection (e.g., Diana, Yonelinas, & Ranganath, 2008), but the only direct test of those competing views—conducted in the context of an associative recognition experiment—supports the continuous recollection interpretation (Mickes et al., 2010). Moreover, Johnson et al. (2009) recently showed that know judgments provided at retrieval are associated with the neural signature of recollection (namely, the recapitulation of encoding-related activity), not with the ab-

sence of recollection-related activity. In addition, as discussed earlier, the process-pure interpretation has great difficulty accommodating the wealth of evidence, reviewed earlier, demonstrating that remember/know judgments, as ordinarily used, reflect different degrees of memory strength.

In the following experiment, words on a list were presented in one of two colors and in one of two locations on the computer screen. The list was followed by a recognition test in which participants were first asked to make an old/new confidence rating for each item using a 20-point scale (Mickes, Wixted, & Wais, 2007). For items declared to be “old,” participants were then asked to make a remember/know/guess judgment and to then recollect source details (item color and screen location). This kind of test has often been performed in the past to test the validity of remember/know judgments. Prior research suggests that remember judgments are associated with higher confidence, higher old/new accuracy, and higher source accuracy than know judgments, but know judgments are also generally associated with above-chance source recollection (not the absence of source recollection). All of these findings are compatible with the standard signal-detection interpretation of remember/know judgments shown in Figures 2 and 3 and with the CDP model shown in Figure 4. Unlike in past research, the focus here is on source accuracy for remember/know judgments equated for high old/new confidence and high old/new accuracy. Is source accuracy for know judgments equal to source accuracy for remember judgments under those conditions? This question has not been previously asked.

## Experiment 1

### Method

**Participants.** Twenty-five undergraduates from University of California, San Diego participated for lower division psychology course credit.<sup>2</sup>

**Materials and design.** The word pool used consisted of 705 three- to seven-letter words taken from the MRC Psycholinguistic Database (Coltheart, 1981), based on a concreteness rating range of 550–700. Of these, 300 words were randomly selected for testing (128 of which were randomly selected to be targets, with another 128 randomly selected to be lures). Instructions and stimuli were displayed for each participant on an LCD monitor and powered by a Dell computer. Stimuli were presented using an E-prime program (Psychology Software Tools; www.psnet.com).

**Procedure.** Participants signed a consent form, were read instructions, studied the 128 targets, and completed a recognition test in which the 128 targets were randomly intermixed with the 128 lures. Each word was presented for 2 s during study. During presentation, words were presented either on top of the screen or on the bottom of the screen and in either red or in blue. The words were balanced in terms of these attributes. Participants were told to memorize the words and their source attributes because memory for the words and the source attributes would be tested. During testing, participants indicated whether or not the word was on the presented list by pressing a number on the keypad ranging from 1 to 20 (with 1 meaning the word was definitely not on the list and with 20 meaning the word was definitely on the list).

After providing a rating using the 20-point confidence scale, participants were asked to make a remember/know/guess judgment

for any item that received a rating of 11 or higher. The instructions for the latter ratings were based on Gardiner and Richardson-Klavehn (2000), and they emphasized that a remember judgment should be made if the source attributes were recalled (i.e., color or location) or if anything else about the presentation of the word was recalled (such as thoughts that the word prompted when it was studied). After the remember/know/guess response was provided, the participant indicated by key press if that word had earlier appeared on the top or bottom of the screen (*t* or *b*, respectively), and then whether the word was originally presented in red or blue (*r* or *b*, respectively). Before studying the list, participants were given a short practice list and a short recognition test to familiarize them with the rating scale, remember/know/guess judgments, and the source-memory test. Any questions that arose after the practice trial were answered before the list was presented.

### Results

The average remember, know, and guess hit rates were 0.24, 0.24 and 0.13, respectively, and the corresponding false alarm rates were 0.03, 0.07, and 0.11. Figure 7A shows the average old/new confidence rating for remember, know, and guess judgments. As is usually true, remember judgments were associated with highest average confidence, know judgments were associated with an intermediate level of confidence, and guesses were associated with a lower level of confidence. Figure 7B shows the average old/new accuracy for remember, know, and guess judgments. Again, as is typically true, accuracy for remember judgments was high, accuracy for know judgments was intermediate, and accuracy for guesses was low. Figure 7B also shows accuracy on the source questions (averaged for color and location, which did not differ from each other), and, again, a graded pattern is observed. Accuracy is highest for remember judgments, next highest for know judgments, and lowest for guesses. For know judgments, source-recollection accuracy was significantly greater than chance, which replicates a pattern that has often been obtained in the past (e.g., Wais et al., 2008). This result indicates that know judgments do not involve the absence of recollection but instead involve lesser degrees of recollection than remember judgments (although it could be argued that the apparent recollection for know judgments instead reflects unitized source familiarity). All of these findings are consistent with the basic signal-detection model of remember/know judgments proposed by Donaldson (1996) and illustrated in Figures 2 and 3. They are also consistent with the CDF model, as illustrated earlier in Figure 4.

We also looked for criterion effects that have been observed in the past. Specifically, we computed the correlation between the remember hit rate and the remember false alarm rate across par-

<sup>2</sup> Five of the 30 participants yielded data that indicated that they were not attentive to the task. One participant skipped many responses, yielding an incomplete data set. Four others exhibited extreme response biases. Three of these participants exhibited an extreme response bias for a rating of 10. The average proportion of target of lures that received a rating of 10 across all 30 participants was 0.16, but for these three participants, the values were 0.55 (and this participant's *d'* was also near chance), 0.77, and 0.81. Finally, one participant exhibited an extreme response bias for a rating of 20 (84% of targets and 79% of lures received that rating). The data from the remaining 25 participants were analyzed.



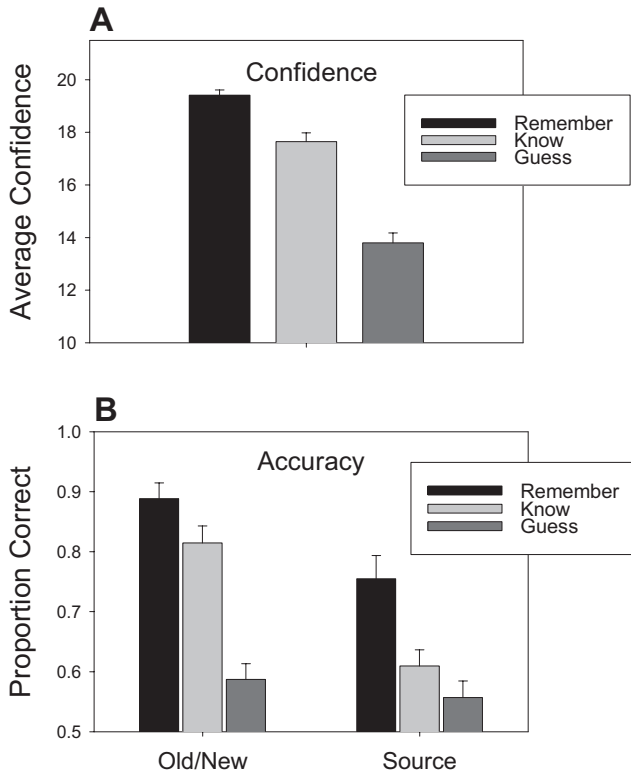


Figure 7. A. Mean confidence associated with remember, know, and guess judgments in Experiment 1. B. Old/new accuracy and source accuracy scores for remember, know, and guess judgments in Experiment 1.

participants. A scatterplot of the 25 scores yielded one extreme outlier. With that individual removed from the analysis, the correlation was .448 ( $p < .05$ ). That is, as is often true, participants with a higher remember hit rate also tended to have a higher false alarm rate, a result that appears to be indicative of a remember criterion. In short, these data are typical in every respect, and they reflect findings that have in the past been taken to demand a signal-detection interpretation of remember/know judgments.

We turned next to the comparison of interest, which is the comparison between remember and know judgments for old decisions that were made with the highest confidence rating of 20 (judgments that are denoted R20 and K20, respectively). The large majority of remember judgments (76%) received that rating, but a substantial proportion of know judgments (37%) did as well. These K20 decisions appear to correspond to the butcher-on-the-bus phenomenon. Of all items rated 20, 61% were remember judgments, 35% were know judgments, and 5% were guesses. Figure 8 shows old/new accuracy associated with remember and know judgments for items rated 20. The accuracy score was approximately 91% for both, and the small difference between them did not approach significance. Thus, these items were, on the surface, equated for strength in terms of both confidence and old/new accuracy. That being the case, it is interesting to ask whether source accuracy was equated for remember and know judgments as well. As shown in Figure 8, source accuracy for R20 judgments was substantially and significantly greater than source accuracy for K20 judgments,  $t(19) = 4.08$ ,  $p < .001$ . This finding suggests that

memories can be strong even if they are based primarily on familiarity, and suggests that people appreciate the difference between strong memories associated with a high degree of recollection and strong memories that involve a relatively weak recollection signal.

The old/new accuracy scores for remember and know judgments made with the highest rating were approximately equal, but they were high enough (both approximately 91% correct) that a difference in strength might have been masked by a ceiling effect. For that reason, these data do not necessarily rule out the criterion variability account. To further investigate this issue, we divided participants on the basis of their old/new accuracy scores associated with remember and know judgments made with the highest level of confidence. That is, for each participant, we computed K20 old/new accuracy minus R20 old/new accuracy. We then divided the participants into two groups, with approximately half consisting of those with the highest K20 – R20 scores (for these 11 participants, average old/new accuracy was higher for K20 than R20) and the other half consisting of those with the lowest K20 – R20 scores (for these 10 participants, average old/new accuracy was lower for K20 than R20). Four participants were not included in this analysis because they did not have any K20 ratings. The main question of interest was whether source accuracy would be higher for R20 than K20 even for participants who had higher old/new K20 scores than R20 scores. Figure 9 shows the results of this analysis. Indeed, even when K20 old/new accuracy exceeded R20 old/new accuracy (see Figure 9A), source-recollection accuracy was higher for items given remember judgments,  $t(10) = 2.45$ . In addition, as both models anticipate, when K20 old/new accuracy was lower than R20 old/new accuracy, source-recollection accuracy was also higher for items given remember judgments (see Figure 9B),  $t(9) = 2.83$ .

Had the source advantage for remember judgments disappeared for the subset of participants with higher K20 old/new scores than R20 old/new scores, the results would have weighed against the CDP model and in favor of a simple strength account. Instead, the results are consistent with the interpretation provided by the CDP model. To reconcile these results with a unidimensional strength model, the assumption would have to be made that memory

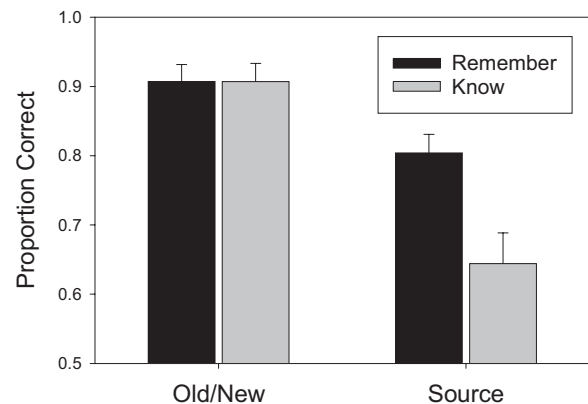


Figure 8. Old/new accuracy and source accuracy scores in Experiment 1 for remember and know judgments associated with items that received the highest confidence (i.e., for items that received a confidence rating of 20).

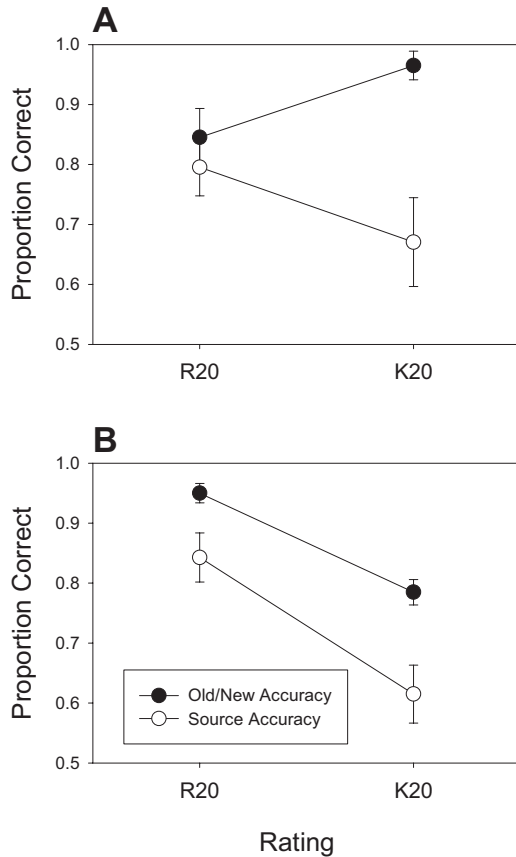


Figure 9. Old/new and source accuracy scores (proportion correct) for R20 and K20 judgments for participants in Experiment 1 whose old/new accuracy was higher for K20 than for R20 (A) and for those whose old/new accuracy was higher for R20 than for K20 (B).

strength for R20 items was higher than that for K20 items but appeared otherwise to be due to random error.

### Experiment 2

Another alternative interpretation of the results of Experiment 1 is that high-confidence know judgments involve more recollection than is apparent in the source accuracy scores. According to this idea, participants sometimes recollect information about the test item that does not involve color or location, such as the thoughts one had about the item when it appeared on the study list. For example, a particular test item, like *beach*, might remind the participant of thoughts that occurred at study about surfing and sunsets, but it might not bring to mind the fact that *beach* appeared in red and at the top of the screen. Even though the instructions clearly stipulated that a remember judgment should be made when recollection of any kind occurred, participants may have been reluctant to choose the remember option in the absence of criterial source recollection (knowing that their criterial source memory was about to be tested). In that case, they might have chosen the know option instead, in which case they would not be expected to know the answer to the source question. If so, K20 memories may have involved just as much recollection as the R20 memories did

(except that the recollection was of a different kind). To test this possibility and to replicate the results of Experiment 1, participants in Experiment 2 were unaware that they would be asked for the source information until after they responded old or new and remember or know or guess to all of the targets and lures.

### Method

**Participants.** Thirty-one undergraduates from University of California, San Diego participated for lower division psychology course credit.

**Materials and design.** These were the same as Experiment 1.

**Procedure.** The procedure was the same as Experiment 1 in every respect except that the source test was presented after participants made old/new and remember/know/guess judgments for targets and lures. The instructions indicated that memory for the words would be tested and that participants should attend to everything about the words, including color and location and any thoughts that came to mind. Although they were told that recognition memory for the words would be tested, no mention was made that memory for source details would also be tested. The recognition test following the short practice list involved old/new confidence ratings and remember/know/guess judgments but no source memory test. The same was true of the recognition test that followed the list proper. The participants were then given the surprise source test on location and color for each of the re-randomized target words.

### Results

Again, the main focus is on remember and know judgments for old decisions that were made with the highest confidence rating (20). The large majority of remember judgments (78%) received that rating, but a substantial proportion of know judgments (50%) did as well. Of all items rated 20, 46% were remember judgments, 51% were know judgments, and 3% were guesses. Figure 10 shows old/new accuracy and source accuracy associated with high-confidence remember and know judgments (i.e., for items rated 20). The old/new accuracy scores were similar for R20 and K20 judgments (94% correct and 91% correct, respectively), but source accuracy for R20 judgments (70% correct) was considerably and significantly higher than source

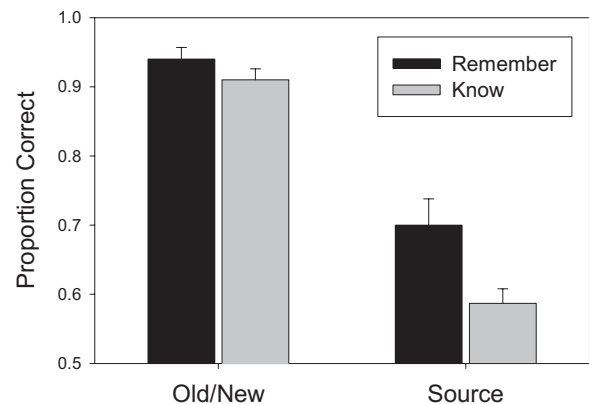


Figure 10. Old/new accuracy and source accuracy scores in Experiment 2 for remember and know judgments associated with items that received the highest confidence (i.e., for items that received a confidence rating of 20).

accuracy for K20 judgments (59%),  $t(27) = 2.81$ . This finding suggests that the source-accuracy difference observed in Experiment 1 did not occur because participants were influenced by the knowledge that their source memory would be tested.

These findings are again consistent with the idea that memories can be strong even if they are based primarily on familiarity. The findings further suggest that people appreciate the difference between strong memories that involve a strong recollection signal and strong memories that involve a relatively weak recollection signal. However, as before, the results could also be explained by arguing that old/new memory strength was not equated. Indeed, in this experiment, old/new accuracy for R20 judgments was slightly higher than old/new accuracy for K20 judgments. That small difference at the high end of the scale could correspond into a large difference at a lower region of the scale (where source-accuracy scores reside). To further investigate this issue, we again divided participants on the basis of their K20 and R20 old/new accuracy scores using the 24 participants who provided scores for both. One group consisted of those with the highest K20 – R20 scores (for these 12 participants, average old/new accuracy was higher for K20 than R20), and the other consisted of those with the lowest K20 – R20 scores (for these 12 participants, average old/new accuracy was lower for K20 than R20). Figure 11 shows the results

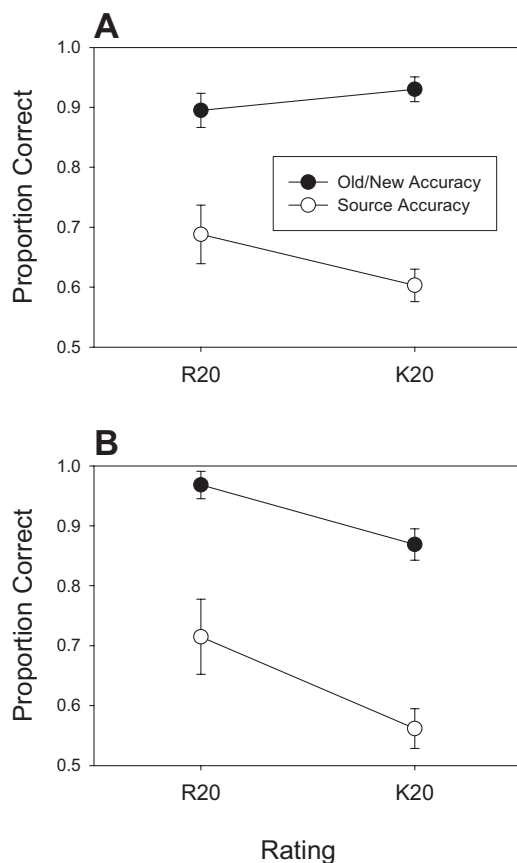


Figure 11. Old/new and source accuracy scores (proportion correct) for R20 and K20 judgments for participants in Experiment 2 whose old/new accuracy was higher for K20 than for R20 (A) and for those whose old/new accuracy was higher for R20 than for K20 (B).

of this analysis. Once again, even when K20 old/new accuracy exceeded R20 old/new accuracy (see Figure 11A), source-recollection accuracy was higher for items given remember judgments, an effect that was marginally significant,  $t(11) = 1.99$ ,  $p = .072$ . Also, as expected by any account, source-recollection accuracy was higher for items given remember judgments when K20 old/new accuracy was lower than R20 old/new accuracy (see Figure 11B),  $t(11) = 2.31$ .

In a further analysis of this issue, we examined old/new accuracy and source accuracy for know judgments associated with confidence ratings of 15 to 20. The question of interest was whether increasing old/new accuracy as confidence increased would be associated with increasing source accuracy. For this analysis, adjacent confidence ratings were collapsed to reduce error variance, and only participants who provided data falling into each (collapsed) cell were included. To maximize power, we combined the data from Experiments 1 and 2, which yielded 34 participants with enough data to analyze. Figure 12 shows that old/new accuracy increased considerably as a function of confidence for know judgments, but source accuracy for know judgments did not vary systematically in the same way. Thus, the data suggest that memories can increase in strength even in the absence of an increase in source memory. Source accuracy was significantly above chance for all three levels of old/new confidence,  $t(33) = 2.09$ , 2.44, and 3.42 for ratings of 15–16, 17–18, and 19–20, respectively, but the mean values were similar (59% correct, 57% correct, and 59% correct, respectively). This result is compatible with the CDP model but is somewhat problematic for a strength model that assumes that source memory strength should increase as old/new memory strength increases (although it could be argued that an increase in source memory was simply not detected because sensitivity to any change in strength is low in that region of the accuracy scale).

### Experiment 3

Experiment 3 was designed to investigate awareness of the content of recognition memory decisions by asking participants for the specifics of what they were remembering. Words were presented on the study list in the same way as the previous two experiments, but the recognition test was different. At test, after making old/new judgments using a 20-point rating scale, participants were provided with five choices: remember color, remember both (location and color), know, and guess. If participants have insight whether they are recollecting color versus location details, it would show up in the respective source accuracy scores. Specifically, the accuracy for color would be greater than the accuracy for location for the items given remember color judgments, and the reverse would be true for items given remember location judgments. This experiment also permitted an interesting test that was not possible to perform in the two previous experiments. In this experiment, unlike the previous ones, many participants claimed to remember source details when their old/new confidence ratings were less than 20. For these participants, we asked how old/new accuracy and source accuracy for relatively low-confidence remember judgments compared with old/new accuracy and source accuracy for high-confidence know judgments (i.e., those associated with a confidence rating of 20).

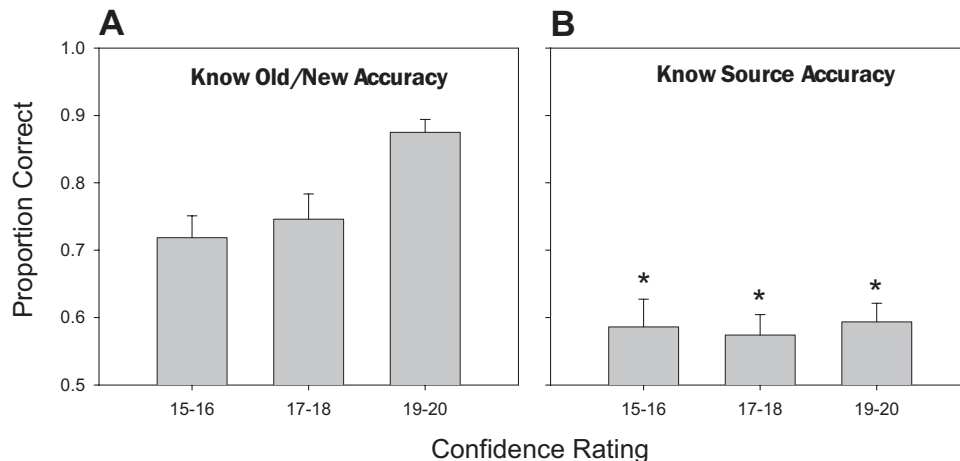


Figure 12. Old/new accuracy (A) and source accuracy (B) for know judgments associated with varying degrees of confidence in the old/new decision. For this analysis, adjacent confidence ratings between 15 and 20 were collapsed, and the data from Experiments 1 and 2 were combined.

## Method

**Participants.** Thirty-five undergraduates from University of California, San Diego participated for lower division psychology course credit.

**Materials and design.** These were the same as the first two experiments.

**Procedure.** The procedure was the same as Experiment 1 in every respect except that after the old/new judgment, five options were presented. The options were remember color, remember location, remember both, know, and guess.

## Results

The first analysis focused on items given an old/new confidence rating of 20. Of the 35 participants, 26 provided judgments that fell into all four of the categories of interest: remember both, remember color, remember source, and know. Figure 13A shows that for these participants, old/new accuracy was highest for remember both (96% correct), somewhat lower for remember color and remember source (90% and 92% correct, respectively), and lowest for know (81% correct). Thus, in this experiment, unlike in the previous ones, know accuracy for items rated 20 was noticeably lower than that for items given a remember judgment of some kind. This finding shows that equating for old/new confidence does not necessarily equate for old/new memory strength. According to the CDP model, this could occur if asking participants to provide details of their source recollection induced them to first check for sufficient recollection and then, if necessary, to check for sufficient familiarity before making old/new confidence ratings for each test item instead of combining recollection and familiarity to make that rating. In that case, separate sets of confidence criteria would be used for remember and know judgments (as illustrated in Figure 5B for a 6-point confidence scale). Under those conditions, equating for confidence in remember and know judgments would not necessarily be expected to equate for accuracy.

One question of interest in this experiment was whether or not participants had access to the details of their source recollection.

Figure 13B shows source accuracy for remember both, remember color, remember source and know judgments. Source accuracy was high for both color and location when participants made remember both judgments, and it was low for both color and location when they made know judgments. Although this is consistent with subjective reports of recollective content, it is also easily explained merely on the basis of old/new strength differences. Of greater interest were the source accuracy scores associated with remember color and remember source judgments. Color accuracy was higher than location accuracy when participants claimed to remember color, and location accuracy was higher than color accuracy when participants claimed to remember location. The interaction between subjective judgment (remember color vs. remember location) and objective performance (color accuracy vs. source accuracy) was significant,  $F(1, 25) = 5.11$ . Thus, these findings suggest that participants have insight into the specific content of their recollective memory when making recognition decisions.

In this experiment, the relative absence of source information also happened to correspond to relatively weak old/new memory (for K-20 judgments). Thus, this finding is easily reconciled with a strength view that holds that know judgments are made when memory is relatively weak. However, participants in this experiment made greater use of the remember options when old/new confidence was less than 20 than did participants in the two previous experiments, particularly when they chose remember color and remember location. This allowed us to compare old/new and source accuracy for relatively weak remember judgments against old/new and source accuracy for relatively strong know judgments. This is similar to the split-half analysis reported for the previous experiments except that, for this analysis, there was no need to divide participants into different subgroups.

For this analysis, we combined data for old/new ratings of 16 to 19 that involved a remember color or a remember location judgment and compared accuracy scores with old/new ratings of 20 that were associated with a know judgment. Figure 14A shows the old/new scores and source accuracy scores (for color and source

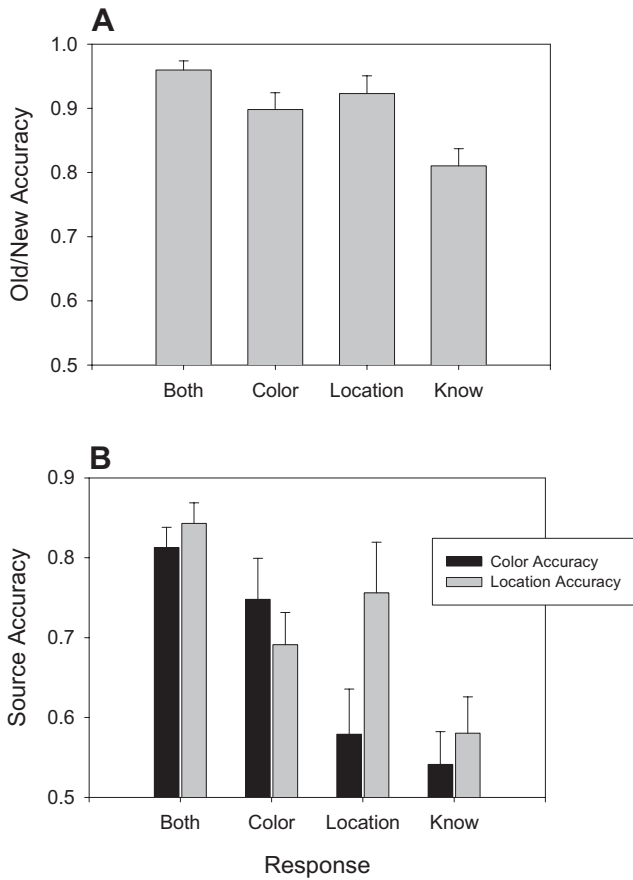


Figure 13. Old/new accuracy scores (A) and source accuracy scores (B) in Experiment 3 for remember both, remember color, remember location, and know judgments associated with items that received the highest confidence (i.e., for items that received a confidence rating of 20).

combined) for the 18 participants who provided analyzable data. The figure shows that old/new accuracy tracks old/new confidence, with K20 accuracy exceeding R16–19 accuracy. However, source accuracy tracks the remember/know judgment such that accuracy is higher for R16–19 than for K20. The interaction was marginally significant,  $t(17) = 1.90, p = .075$ . Although this result is not definitive, it is suggestive.

We next computed old/new and source accuracy scores for combined confidence ratings of 16 to 19 that were associated with a know judgment (K16–19). With these additional scores (along with the data shown in Figure 14A), a state-trace plot could be constructed (Dunn, 2008). The value of a state-trace plot is that it provides a more definitive way to test whether remember and know decisions are based on a unidimensional memory-strength variable. If so, the points should all trace out as a continuous, monotonically increasing function. If more than one dimension is involved, then the function for know judgments should be distinct from the function for remember judgments. Figure 14B shows the state-trace plot for this experiment based on data from 15 participants who supplied enough data in all four categories to perform this analysis. The functions appear to differ (i.e., the points do not follow a monotonically increasing function), a result that is consistent with a multidimensional account.

The results from the three experiments are easily reconciled with the CDP model, and this would not be true had the results turned out otherwise. Thus, in light of these findings, and in light of the model’s ability to accommodate a wealth of prior evidence that has proven to be problematic for other dual-process accounts of remember/know judgments, we argue that the CDP model is a viable dual-process competitor to the currently dominant unidimensional signal detection account (unlike other dual-process accounts). The findings can also be reconciled with the unidimensional signal detection account, but it seems fair to say that the results could have turned out differently in every case and in a way that would have been even more easily reconciled with that account. Even so, our point is not that the CDP model has been established to be superior to the simpler unidimensional account on the basis of these findings. Instead, our point is that the results establish the viability of the CDP model. Because it is the only dual-process theory that can naturally accommodate prior remember/know studies supporting a strength-based interpretation, the CDP model is a more viable contender against the dominant unidimensional signal detection account than other dual-process models.

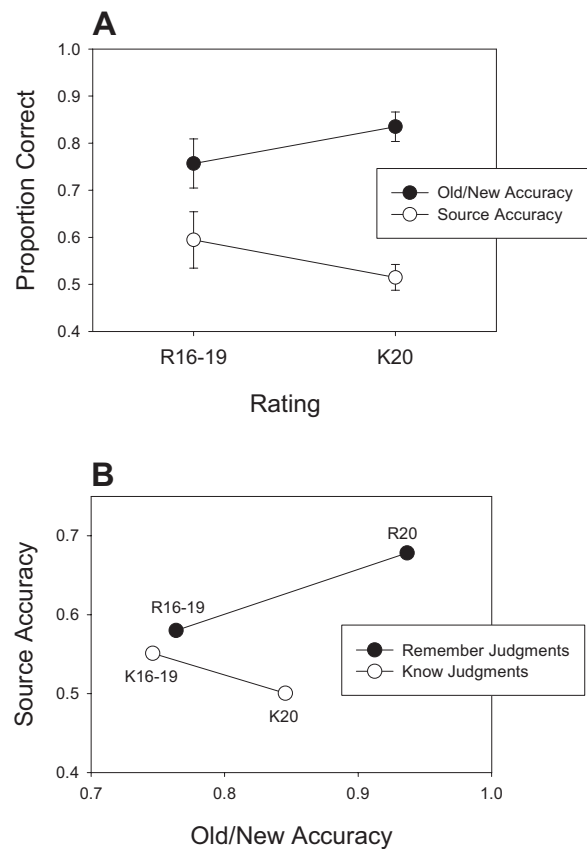


Figure 14. A. Old/new and source accuracy scores (proportion correct) associated with confidence ratings of 16 through 19 combined and followed by remember judgments (R16–19) and confidence ratings of 20 and followed by know judgments (K20) for participants in Experiment 3. B. State-trace plot for old/new and source accuracy scores associated with two levels of confidence (16–19 vs. 20) for know (K) and remember (R) judgments.

Finally, the results summarized in Figure 14 point the way toward the kind of test that could help to more definitively differentiate between competing accounts. The observed interaction (see Figure 14A) and the obtained state-trace plot (see Figure 14B)—should they be more firmly established in future work—are not as easily reconciled with the unidimensional signal-detection model as our other findings are. In addition, the very existence of remember judgments made with lower levels of confidence and lower levels of old/new accuracy is hard to reconcile with the DPSD model (which does not view recollection as a continuous process that can yield different levels of confidence). Moreover, the fact that the R16–R19 judgments were associated with considerable source recollection cannot be easily explained by appealing to “unitized familiarity” or to any other kind of familiarity, because participants declared that recollection was involved (i.e., they supplied remember judgments), and their source accuracy scores confirm that this was the case. Experimental tests of this sort (namely, comparing recollection associated with low-confidence remember judgments vs. high-confidence know judgments) hold the promise of effectively differentiating between unidimensional signal-detection-based accounts, extant dual-process accounts such as the DPSD model, and the new CDP model.

### General Discussion

The concept of memory strength has a long and useful history both because it provides an intuitive interpretation of behavioral variables that tend to covary (namely, confidence, accuracy and reaction time) and because it lends itself to more formal specification in terms of signal detection theory (which, in turn, helps to conceptualize a variety of memory-related phenomena). However, its long-appreciated weakness is that any characterization of memory in terms of strength seems to deny the characterization of memory in terms of content, and it seems clear that memory—including recognition memory—is rich in content. These considerations have, in recent years, been particularly evident in the debate over the relative validity of the signal-detection versus dual-process interpretations of remember/know judgments. In this debate, a strength-based view (signal-detection theory) has been pitted against a largely content-based view (dual-process theory). However, our central claim is that these are not inherently incompatible points of view. Indeed, the proposed CDP model illustrated in Figure 4 not only respects both dual-process theory and signal-detection theory (as much evidence would seem to demand) but it also recognizes the validity of the widely used remember/know procedure—but only in modified form. Indeed, if the CDP model is valid, the conclusions of many prior studies that have used the remember/know procedure to investigate recollection and familiarity (including virtually all neuroimaging studies that have used that procedure) may not be valid because, as ordinarily used, remember and know judgments merely reflect strong and weak memories, not recollection and familiarity.

The CDP model holds that remember judgments ordinarily will be made with higher confidence and higher accuracy than know judgments, in agreement with the standard signal-detection account of remember/know judgments illustrated in Figure 2. However, it avoids an awkward implication of the standard signal-detection view, which is that strong, familiarity-based decisions either do not occur (despite compelling subjective experiences that

suggest otherwise) or that, if they do occur, participants will supply a remember judgment (as if they do not realize that the decision was based on strong familiarity). The CDP model illustrated in Figure 4, by contrast, assumes that strong, familiarity-based decisions do occasionally occur, that participants know when they are experiencing strong memory in the absence of strong recollection, and that they appropriately provide a know judgment under those circumstances. However, even in that case, a know judgment does not imply the absence of recollection (i.e., it does not signify a pure familiarity-based experience). Instead, it implies that the strength of recollection did not exceed the participant’s criterion for declaring the item to have been remembered and that the recognition decision is primarily based on strong familiarity.

### The Concept of “Memory Strength”

A key theoretical consideration in the CDP model is that “memory strength” is not a unitary construct, which means that it is not properly conceptualized as a fixed property of the encoded memory trace independent of the conditions of retrieval. If memory strength is construed as a property of the trace itself (as it sometimes is), then certain well-known contradictions arise. For example, it has long been known that recognition memory is better for low-frequency words than for high-frequency words, but when pure-frequency lists are used, the reverse is true for recall. If low-frequency words are associated with stronger memory than high-frequency words, why would this reversal ever be observed?

In a recent article, Cohn, Moscovitch, Lahat, and McAndrews (2009) stated that “Proponents of the strength view place their emphasis on the encoded trace, leaving the question of ecphoric strength (i.e., combination of retrieval cues and activated information) somewhat ambiguous in their accounts” (p. 22454). However, this statement does not correspond to what proponents of the strength view believe. In the CDP model, strength is not solely a property of the memory trace. Instead, strength is a property of the *signal* that is returned in response to a retrieval cue. As Tulving has repeatedly and convincingly demonstrated, one’s experience of memory is a joint function of the encoded trace and the extant retrieval cues (e.g., Tulving & Thompson, 1972). As such, one cannot speak of the strength of memory at a particular moment in time without also taking into account the way in which memory was queried at that moment in time. A similar point has been made in connection with the notion of transfer-appropriate processing, which holds that a retrieval cue is effective to the extent that it replicates the operations that were performed at the time of encoding (Roediger, Weldon, & Challis, 1989).

The CDP model envisions a variety of different memory-strength variables, not just one. When memory is queried by asking whether the test item is old or new, the memory-strength variable is assumed to consist of the combination of the continuous recollection and familiarity signals (because, under those conditions, there may be no reason for the participant to treat them separately). When the participant is next asked to make a remember/know judgment, the operative memory-strength variable is determined by two separate queries of memory. The first occurs when the participant asks, “Am I experiencing enough recollection to declare that I remember this item?” According to the CDP model, if the strength of the recollection signal that is returned in

response to this query exceeds a criterion value, a remember judgment is made. If not, memory is queried again and in another way. This occurs when the participant asks, “Am I experiencing enough familiarity to declare that I know that this item was on the list?” If the strength of the familiarity signal that occurs in response to this query exceeds a criterion value, a know judgment is made (otherwise, the guess option is chosen). Still another memory signal is returned when memory is queried yet again by a source memory question, such as, “Was this item presented at the top of the screen or the bottom of the screen?”

The critical point is that memory strength is not a unitary construct (cf. Banks, 2000). Instead, it is a *signal* that is returned by the memory system and that depends, in part, on the way in which memory is queried (cf. Humphreys et al., 2003). An additional consideration—one that is just as critical—is the fact that no matter how memory is queried, a memory signal is always returned, even if the test item is a lure. The recollection signal returned by a lure may generally be weak unless steps are taken to make it strong, such as in the DRM procedure (Roediger & McDermott, 1995), but the essence of a signal-detection model is that a memory signal is returned, even if the retrieval cue involves a test item that was not present at encoding, and even if that lure was not indirectly activated at encoding. The fact that a signal is returned for both targets and lures defines the signal-detection problem faced by the participant (and it is why a decision criterion is needed).

### Signal-Detection Theory and the Mechanisms of Memory

The CDP signal-detection model is a model about how participants base decisions on recollection and familiarity signals that occur at the time of retrieval. It says nothing about the mechanisms that give rise to those signals. That is, instead of explaining the recollection and familiarity signals, it assumes them. It is conceivable that a mechanistic model involving only one memory process will be able to accommodate high-confidence recognition decisions that involve associative detail (which we refer to as recollection-based decisions) as well as high-confidence recognition decisions that do not (which we refer to as familiarity-based decisions). However, we assume that two processes are involved because, in our view, making that assumption is the most parsimonious way to account for our remember/know data.

A computational model that does speak to the mechanisms that underlie recollection and familiarity and that also has some similarities to the CDP model is the source of activation confusion (SAC) model advanced by Reder et al. (2000). This model envisions *episode nodes* and *concept nodes*, both of which can be activated to varying degrees to yield experiences that are akin to recollection and familiarity, respectively. The concept node represents information about an item that has been stored from previous experience with that item, so it can be activated by targets and lures (and its activation determines familiarity). The episode node represents the experience of studying the word in the experimental context, so it is activated by targets only (and its activation determines recollection).

If activation in the episode node exceeds what Reder et al. (2000) refer to as a “threshold” (but which may correspond to what we would call an adjustable criterion), a remember judgment is

made. If not, then if activation in the concept node exceeds a threshold (which, again, might be better termed a criterion), a know judgment is made. This is much like the decision process envisioned by the CDP model for making remember/know judgments. A critical difference is that the CDP model envisions target and lure distributions for both recollection and familiarity (as shown in Figure 4 and Figure 5B). In SAC, there is no lure distribution for recollection because lures, having not appeared on the list, do not have an episode node. Note that, to a large extent, the existence (or not) of a meaningful lure distribution for recollection is the essence of the difference between a signal-detection interpretation of recollection and a threshold view of recollection.

Because SAC does not envision a lure distribution for recollection, the existence of remember false alarms, as well as the commonly observed correlation between remember hit and false alarm rates (Wixted & Stretch, 2004), is problematic for this model. Diana et al. (2006) suggested that SAC may need to be modified to allow for the possibility of false recollection of the features of lures (features that were also part of targets and so are connected to an episode node). This would come very close to introducing a lure distribution for recollection. If the threshold on the episode node is also construed as a decision criterion, one that could differ in its placement for different participants, then SAC could also account for the typical correlation between remember hit and false alarm rates. Thus, despite their differences, it seems clear that SAC offers the most natural marriage between the CDP signal-detection decision model shown in Figure 4 and a mechanistic model that specifies how the recollection and familiarity signals arise.

### Practical Implications for Using the Remember/Know Procedure

If the CDP signal-detection model offers an accurate way to conceptualize recognition memory, then using the remember/know procedure without taking into account confidence and accuracy (as is typically the case) is not an effective way to distinguish between recollection and familiarity. The reason is that, if confidence is not equated, remember judgments reflect strong memory (not recollection), whereas know judgments reflect weaker memory (not familiarity). For example, under typical conditions, tests of source memory show that know judgments are generally associated with lesser degrees of recollection than are remember judgments, not with the absence of recollection (Johnson et al., 2009; Wais et al., 2008). Moreover, if recollection and familiarity are correlated processes, as seems likely (e.g., Humphreys et al., 2003), then know judgments might also be associated with lesser degrees of familiarity compared to remember judgments. To address these issues, old/new memory strength associated with remember and know judgments should at least be equated for confidence, a point made by Rotello and Zeng (2008) in their investigation in reaction times associated with remember and know judgments. As the large majority of remember judgments are associated with high confidence, this may often entail using only high-confidence old/new decisions, as we did in Experiments 1 and 2.

Although the recommendation to equate confidence derives from the CDP signal-detection model proposed in Figure 4, the selective use of high-confidence know judgments (vs. high-confidence remember judgments) should not be viewed as prob-

lematic from any dual-process perspective. In fact, standard remember/know instructions clearly stipulate that know judgments should be given only when the participant has high confidence that the item appeared on the list. Consider, for example, the instructions for know judgments from Rajaram (1993), which are widely used:

“Know” responses should be made when . . . you are *certain* of recognizing the words but these words fail to evoke any specific conscious recollection from the study list. [emphasis added]

In practice, participants generally do not heed this instruction and instead tend to supply know judgments when they are not certain of their decision. Thus, it makes sense to take steps to ensure that know judgments are associated with high confidence (and high accuracy).

Although equating for old/new confidence is an important step in equating for memory strength, even that may not be enough. For example, even if recollection and familiarity are ordinarily summed to make old/new decisions (as the CDP model assumes), it seems possible that requesting remember/know decisions on every test trial might induce participants to sequentially check recollection and then familiarity before making an old/new decision. In that case, two sets of confidence criteria would be involved: one on the recollection axis and the other on the familiarity axis (see Dunn, 2008, for a similar suggestion in the context of a single process account). That is, the confidence criteria would not be placed on the aggregate memory-strength axis (as in Figure 4) but would instead be placed on both the recollection and familiarity signal-detection models (as in Figure 5B). Under those conditions, equating for confidence would not necessarily equate for strength (as was the case for ratings of 20 in our Experiment 3). Equating for old/new accuracy in addition to equating for old/new confidence is helpful in this regard.

In the experiments presented earlier, we analyzed a subset of know judgments (i.e., those associated with confidence ratings of 20) in an effort to equate for strength, but other approaches have been used previously. For example, neuroimaging studies sometimes make use of a variant of the remember/know procedure by asking participants to use a 1–2–3–4–R confidence scale, where 1 through 4 denote varying degrees of confidence that the item is new or old (1 = *sure new*, 2 = *maybe new*, 3 = *maybe old*, and 4 = *sure old*) and R denotes decisions based on recollection (e.g., Cohn et al., 2009; Yonelinas, Otten, Shaw, & Rugg, 2005). If recollection is a threshold process, whereas familiarity is a continuous process, as the DPSD model assumes, then ratings of 1 through 4 represent varying degrees of familiarity (and they are effectively know judgments), whereas ratings of R represent the occurrence of categorical recollection. Thus, ratings of 4 should represent high-confidence familiarity-based decisions.

Cohn et al. (2009) went so far as to instruct participants that ratings of 4 and R should differ in terms of the presence or absence of recollection but should not differ in confidence or strength. However, they did not compute old/new accuracy scores for those ratings to test whether strength was at least approximately equated, but this can be done using the hit and false alarm rate data presented in their Table 2. In one of their conditions (the “uncued target” condition), old/new accuracy was 0.81 for ratings of 4 and 0.95 for ratings of R. In another condition (the “cued target” condition), old/new accuracy was 0.76 for ratings of 4 and 0.98 for

ratings of R. Thus, as in our Experiment 3, it is clear that this scale did not succeed in equating for memory strength for high-confidence remember and know judgments. As such, finding that activity in the hippocampus is elevated for ratings of R but not for ratings of 4 could mean that the hippocampus selectively subserves recollection (a common interpretation) or it could instead mean that elevated hippocampal activity is hard to detect unless memory is strong (cf. Squire, Wixted, & Clark, 2007).

The key point is that equating for strength when comparing remember/know judgments is not trivially easy. As such, it would make sense to design scales that structurally attempt to equate for strength and to then test those scales to determine whether the remember and know judgments of interest have similar old/new accuracy but different source accuracy. An example might be a scale such as 1–2–3–4–5 (1 = *sure new*, 2 = *probably new*, 3 = *maybe new*, 4 = *maybe old*, 5 = *probably old*) followed by two high-confidence options, 6R or 6K (6R = *sure old remember* and 6K = *sure old know*). This would procedurally instantiate what the instructions have long asked participants to do and which many have mistakenly assumed that participants are doing in the large and ever growing number of experiments that make use of the remember/know procedure. Another possible rating scale variant would be to have old decisions rated as 4R/4K, 5R/5K, 6R/6K. Whatever scale is used, it would be prudent to calculate old/new accuracy for remember and know judgments to see if accuracy has been equated in addition to confidence. If not, conclusions should probably be tempered accordingly.

## Conclusion

The concept of memory strength—and the idea that lures as well as targets yield a memory-strength signal (including a recollection signal)—seems like an essential ingredient of any theory of recognition memory. The question of whether the memory-strength signal should be construed in terms of a signal-process signal-detection model (as is ordinarily assumed by signal-detection theorists) or as a dual-process signal-detection model (as illustrated in Figure 4) hinges to a large extent on the existence (or not) of the butcher-on-the-bus phenomenon. A single-process signal-detection model has trouble accommodating high-confidence familiarity-based decisions made with high accuracy, whereas a dual-process signal-detection model does not. Subjective experience, and the evidence presented in our Experiments 1 and 2, suggests that the butcher-on-the-bus experience is real and must somehow be explained. The CDP model illustrated in Figure 4 is unique in that it accounts for the butcher-on-the-bus phenomenon while also accounting for the large body of evidence that demands a strength-based interpretation of remember/know judgments (as that procedure is ordinarily used). Further empirical inquiry into the nature of the butcher-on-the-bus phenomenon should help to more convincingly differentiate competing single-process and dual-process accounts of recognition memory.

## References

- Aizpurua, A., Garcia-Bajos, E., & Migueles, M. (2009). Memory for actions of an event: Older and younger adults compared. *Journal of General Psychology*, *136*, 428–441.



- Anderson, J. R., & Bower, G. H. (1972). Recognition and retrieval processes in free recall. *Psychological Review*, *79*, 97–123.
- 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.
- Atkinson, R. C., & Juola, J. F. (1974). Search and decision processes in recognition memory. In D. H. Krantz, R. C. Atkinson, & P. Suppes (Eds.), *Contemporary developments in mathematical psychology* (pp. 243–290). San Francisco, CA: Freeman.
- Banks, W. P. (2000). Recognition and source memory as multivariate decision processes. *Psychological Science*, *11*, 267–273.
- Benjamin, A. S., Diaz, M. L., & Wee, S. (2009). Signal detection with criterion noise: Applications to recognition memory. *Psychological Review*, *116*, 84–114.
- Carr, V. A., Viskontas, I. V., Engel, S. A., & Knowlton, B. (2009). Neural activity in the hippocampus and perirhinal cortex during encoding is associated with the durability of episodic memory. *Journal of Cognitive Neuroscience*, *22*, 2652–2662.
- Clarys, D., Bugajska, A., Tapia, G., & Baudouin, A. (2009). Ageing, remembering, and executive function. *Memory*, *17*, 158–168.
- Cohen, A. L., Rotello, C. M., & Macmillan, N. A. (2008). Evaluating models of remember-know judgments: Complexity, mimicry, and discriminability. *Psychonomic Bulletin & Review*, *15*, 906–926.
- Cohn, M., Moscovitch, M., Lahat, A., & McAndrews, M. P. (2009). Recollection versus strength as the primary determinant of hippocampal engagement at retrieval. *Proceedings of the National Academy of Sciences, USA*, *106*, 22451–22455.
- Coltheart, M. (1981). The MRC Psycholinguistic Database. *Quarterly Journal of Experimental Psychology*, *33*, 497–505.
- Daury, N. (2009). Gaze direction influences awareness in recognition memory for faces after intentional learning. *Perceptual and Motor Skills*, *109*, 224–234.
- DeCarlo, L. T. (2003). Source monitoring and multivariate signal detection theory, with a model for selection. *Journal of Mathematical Psychology*, *47*, 292–303.
- de Chastelaine, M., Friedman, D., Cycowicz, Y. M., & Horton, C. (2009). Effects of multiple study-test repetition on the neural correlates of recognition memory: ERPs dissociate remembering and knowing. *Psychophysiology*, *46*, 86–99.
- Dewhurst, S. A., Conway, M. A., & Brandt, K. R. (2009). Tracking the R-to-K shift: Changes in memory awareness across repeated tests. *Applied Cognitive Psychology*, *23*, 849–858.
- Dewhurst, S. A., Holmes, S. J., Brandt, K. R., & Dean, G. M. (2006). Measuring the speed of the conscious components of recognition memory: Remembering is faster than knowing. *Consciousness and Cognition*, *15*, 147–162.
- Diana, R., Reder, L. M., Arndt, J., & Park, H. (2006). Models of recognition: A review of arguments in favor of a dual process account. *Psychonomic Bulletin & Review*, *13*, 1–21.
- Diana, R. A., Yonelinas, A. P., & Ranganath, C. (2008). The effects of unitization on familiarity-based source memory: Testing a behavioral prediction derived from neuroimaging data. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 730–740.
- Dobbins, I. G., Khoe, W., Yonelinas, A. P., & Kroll, N. E. A. (2000). Predicting individual false alarm rates and signal detection theory: A role for remembering. *Memory & Cognition*, *28*, 1347–1356.
- Donaldson, W. (1996). The role of decision processes in remembering and knowing. *Memory & Cognition*, *24*, 523–533.
- Dorfel, D., Wener, A., Schaefer, M., von Kummer, R., & Karl, A. (2009). Distinct brain networks in recognition memory share a defined region in the precuneus. *European Journal of Neuroscience*, *30*, 1947–1959.
- Dougal, S., & Rotello, C. M. (2007). “Remembering” emotional words is based on response bias, not recollection. *Psychonomic Bulletin & Review*, *14*, 423–429.
- Duarte, A., Henson, R. N., & Graham, K. S. (2008). The effects of aging on the neural correlates of subjective and objective recollection. *Cerebral Cortex*, *18*, 2169–2180.
- Dunn, J. C. (2004). Remember-know: A matter of confidence. *Psychological Review*, *111*, 524–542.
- Dunn, J. C. (2008). The dimensionality of the remember-know task: A state-trace analysis. *Psychological Review*, *115*, 426–446.
- Egan, J. P. (1958). *Recognition memory and the operating characteristic* (Tech. Note AFCRC-TN-58–51). Bloomington, IN: Indiana University, Hearing and Communication Laboratory.
- Eldridge, L. L., Engel, S. A., Zeineh, M. M., Bookheimer, S. Y., & Knowlton, B. J. (2005). A dissociation of encoding and retrieval processes in the human hippocampus. *Journal of Neuroscience*, *25*, 3280–3286.
- Friedman, D., de Chastelaine, M., Nessler, D., & Malcolm, B. (2009, November 13). Changes in familiarity and recollection across the lifespan: An ERP perspective. *Brain Research*, *1310*, 124–141.
- Gardiner, J. M. (1988). Functional aspects of recollective experience. *Memory & Cognition*, *16*, 309–313.
- Gardiner, J., Gawlik, B., & Richardson-Klavehn, A. (1994). Maintenance rehearsal affects knowing, not remembering; elaborative rehearsal affects remembering, not knowing. *Psychonomic Bulletin & Review*, *1*, 107–110.
- Gardiner, J. M., & Java, R. I. (1990). Recollective experience in word and nonword judgments. *Memory & Cognition*, *18*, 23–30.
- Gardiner, J., & Richardson-Klavehn, A. (2000). Remembering and knowing. In E. Tulving & F. I. M. Craik (Eds.), *The Oxford handbook of memory*. New York, NY: Oxford University Press.
- Gardiner, J., 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.
- Geraci, L., McCabe, D. P., & Guillory, J. J. (2009). On interpreting the relationship between remember-know judgments and confidence: The role of instructions. *Consciousness and Cognition*, *18*, 701–709.
- Glanzer, M., Adams, J. K., Iverson, G. J., & Kim, K. (1993). The regularities of recognition memory. *Psychological Review*, *100*, 546–567.
- Gomez, A., Rousset, S., & Baciú, M. (2009). Egocentric-updating during navigation facilitates episodic memory retrieval. *Acta Psychologica*, *132*, 221–227.
- Harkin, B., & Kessler, K. (2009). How checking breeds doubt: Reduced performance in a simple working memory task. *Behaviour Research and Therapy*, *47*, 504–512.
- Hautus, M., Macmillan, N. A., & Rotello, C. M. (2008). Toward a complete decision model of item and source recognition. *Psychonomic Bulletin & Review*, *15*, 889–905.
- Hess, T. M., Emery, L., & Queen, T. L. (2009). Task demands moderate stereotype threat effects on memory performance. *The Journals of Gerontology, Series B: Psychological Sciences and Social Sciences*, *64*, 482–486.
- Hicks, J. L., Marsh, R. L., & Ritschel, L. (2002). The role of recollection and partial information in source monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 503–508.
- Hirshman, E., & Henzler, A. (1998). The role of decision processes in conscious recollection. *Psychological Science*, *9*, 61–65.
- Hirshman, E., & Master, S. (1997). Modeling the conscious correlates of recognition memory: Reflections on the remember-know paradigm. *Memory & Cognition*, *25*, 345–351.
- Hockley, W. E., & Cristi, C. (1996). Tests of encoding tradeoffs between item and associative information. *Memory & Cognition*, *24*, 202–216.
- Hudon, C., Belleville, S., & Gauthier, S. (2009). The assessment of recognition memory using the remember/know procedure in amnesic mild cognitive impairment and probable Alzheimer’s disease. *Brain and Cognition*, *70*, 171–179.
- Humphreys, M., Dennis, S., Maguire, A., Reynolds, K., Bolland, S., &

- Hughes, D. (2003). What you get out of memory depends on the question you ask. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *29*, 797–812.
- Inoue, C., & Bellezza, F. S. (1998). The detection model of recognition using know and remember judgments. *Memory & Cognition*, *26*, 299–308.
- Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language*, *30*, 513–541.
- 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.
- Jermann, F., Van der Linden, M., Laurencon, M., & Schmitt, B. (2009). Recollective experience during recognition of emotional words in clinical depression. *Journal of Psychopathology and Behavioral Assessment*, *31*, 27–35.
- Johnson, J. D., McDuff, S. G. R., Rugg, M. D., & Norman, K. A. (2009). Recollection, familiarity, and cortical reinstatement: A multivoxel pattern analysis. *Neuron*, *63*, 697–708.
- Klumpp, H., Amir, N., & Garfinkel, S. N. (2009). False memory and obsessive-compulsive symptoms. *Depression and Anxiety*, *26*, 396–402.
- Kurilla, B. P., & Westerman, D. L. (2010). Source memory for unidentified stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*, 398–410.
- Lemogne, C., Bergouignan, L., Piolino, P., Jouvent, R., Allilaire, J.-F., & Fossati, P. (2009). Cognitive avoidance of intrusive memories and autobiographical memory: Specificity, autoevident consciousness, and self-perspective. *Memory*, *17*, 1–7.
- MacLaverly, S. N., & Hertzog, C. (2009). Do age-related differences in episodic feeling of knowing accuracy depend on the timing of the judgment? *Memory*, *17*, 860–873.
- Mandler, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological Review*, *87*, 252–271.
- McCabe, D. P., & Geraci, L. D. (2009a). The influence of instructions and terminology on the accuracy of remember-know judgments. *Consciousness and Cognition*, *18*, 401–413.
- McCabe, D. P., & Geraci, L. D. (2009b). The role of extralist associations in false remembering: A source misattribution account. *Memory & Cognition*, *37*, 130–142.
- Mickes, L., Johnson, E., & Wixted, J. T. (2010). Continuous recollection vs. unitized familiarity in associative recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*, 843–863.
- 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.
- Mickes, L., Wixted, J. T., & Wais, P. E. (2007). A direct test of the unequal-variance signal-detection model of recognition memory. *Psychonomic Bulletin & Review*, *14*, 858–865.
- Migo, E., Montaldi, D., Norman, K. A., Quamme, J., & Mayes, A. (2009). The contribution of familiarity to recognition memory is a function of test format when using similar foils. *Quarterly Journal of Experimental Psychology*, *62*, 1198–1215.
- Norman, D. A., & Wickelgren, W. A. (1969). Strength theory of decision rules and latency in retrieval from short-term memory. *Journal of Mathematical Psychology*, *6*, 192–208.
- Onyper, S. V., Zhang, Y., & Howard, M. W. (2010). Some-or-none recollection: Evidence from item and source memory. *Journal of Experimental Psychology: General*, *149*, 341–364.
- Palmer, M. A., Brewer, N., McKinnon, A. C., & Weber, N. (2010). Phenomenological reports diagnose accuracy of eyewitness identification decisions. *Acta Psychologica*, *133*, 137–145.
- Parker, A., Buckley, S., & Dagnall, N. (2008). Reduced misinformation effects following saccadic bilateral eye movements. *Brain and Cognition*, *69*, 89–97.
- Parker, A., & Dagnall, N. (2009). Concreteness effects revisited: The influence of dynamic visual noise on memory for concrete and abstract words. *Memory*, *17*, 397–410.
- Parks, C. M., & Yonelinas, A. P. (2007). Moving beyond pure signal-detection models: Comment on Wixted. *Psychological Review*, *114*, 188–202.
- Parks, C. M., & Yonelinas, A. P. (2009). Evidence for a memory threshold in second-choice recognition memory responses. *Proceedings of the National Academy of Sciences USA*, *106*, 11515–11519.
- Peker, M., & Tekcan, A. I. (2009). The role of familiarity among group members in collaborative inhibition and social contagion. *Social Psychology*, *40*, 111–118.
- Rajaram, S. (1993). Remembering and knowing: Two means of access to the personal past. *Memory & Cognition*, *2*, 89–102.
- Ratcliff, R., & Murdock, B. B., Jr. (1976). Retrieval processes in recognition memory. *Psychological Review*, *83*, 190–214.
- Ratcliff, R., & Starns, J. J. (2009). Modeling confidence and response time in recognition memory. *Psychological Review*, *116*, 59–83.
- 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.
- Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 803–814.
- Roediger, H. L., Weldon, M. S., & Challis, B. H. (1989). Explaining dissociations between implicit and explicit measures of retention: A processing account. In H. L. Roediger & F. I. M. Craik (Eds.), *Varieties of memory and consciousness: Essays in honour of Endel Tulving* (pp. 3–39). Hillsdale, NJ: Erlbaum.
- 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.
- 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.
- 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.
- Rotello, C. M., & Zeng, M. (2008). Analysis of RT distributions in the remember-know paradigm. *Psychonomic Bulletin & Review*, *15*, 825–832.
- Sauerland, M., & Sporer, S. L. (2009). Fast and confident: Postdicting eyewitness identification accuracy in a field study. *Journal of Experimental Psychology: Applied*, *15*, 46–62.
- Sherman, S. J., Atri, A., Hasselmo, M. E., Stern, C. E., & Howard, M. W. (2003). Scopolamine impairs human recognition memory: Data and modeling. *Behavioral Neuroscience*, *117*, 526–539.
- Shimamura, A. P., & Wickens, T. D. (2009). Superadditive memory strength for item and source recognition: The role of hierarchical relational binding in the medial temporal lobe. *Psychological Review*, *116*, 1–19.
- Skinner, E. I., & Fernandes, M. A. (2009). Illusory recollection in older adults and younger adults under divided attention. *Psychology and Aging*, *24*, 211–216.
- Slotnick, S. D. (2010). “Remember” source memory ROCs indicate recollection is a continuous process. *Memory*, *18*, 27–39.
- Slotnick, S. D., & Dodson, C. S. (2005). Support for a continuous (single-process) model of recognition memory and source memory. *Memory & Cognition*, *33*, 151–170.
- Squire, L. R., Wixted, J. T., & Clark, R. E. (2007). Recognition memory

- and the medial temporal lobe: A new perspective. *Nature Reviews Neuroscience*, 8, 872–883.
- Starns, J. J., Hicks, J. L., Brown, N. L., & Martin, B. A. (2008). Source memory for unrecognized items: Predictions from multivariate signal detection theory. *Memory & Cognition*, 36, 1–8.
- 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.
- Stoettinger, E., Kaiser, W., & Perner, J. (2009). Remember judgments and the constraint of direct experience. *Psychological Research*, 73, 623–632.
- 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.
- Tanweer, T., Rathbone, C. J., & Souchay, C. (2009, November 13). Autobiographical memory, autoegetic consciousness, and identity in Asperger syndrome. *Neuropsychologia*, 48, 900–908.
- Tulving, E. (1985). Memory and consciousness. *Canadian Psychology*, 26, 1–12.
- Tulving, E., & Thompson, D. M. (1972). Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, 80, 352–373.
- Vilberg, K. L., & Rugg, M. D. (2009a). Functional significance of retrieval-related activity in lateral parietal cortex: Evidence from fMRI and ERPs. *Human Brain Mapping*, 30, 1490–1501.
- Vilberg, K. L., & Rugg, M. D. (2009b). Left parietal cortex is modulated by amount of recollected verbal information. *NeuroReport*, 20, 1295–1299.
- Viskontas, I. V., Carr, V. A., Engel, S. A., & Knowlton, B. J. (2009). The neural correlates of recollection: Hippocampal activation declines as episodic memory fades. *Hippocampus*, 19, 265–272.
- Voss, J. L., & Paller, K. A. (2009). Remembering and knowing: Electrophysiological distinctions at encoding but not retrieval. *NeuroImage*, 46, 280–289.
- Wais, P. E., Mickes, L., & Wixted, J. T. (2008). Remember/know judgments probe degrees of recollection. *Journal of Cognitive Neuroscience*, 20, 400–405.
- Wheeler, M. E., & Buckner, R. L. (2004). Functional-anatomic correlates of remembering and knowing. *NeuroImage*, 21, 1337–1349.
- Wiesmann, M., & Ishai, A. (2008). Recollection- and familiarity-based decisions reflect memory strength. *Frontiers in Systems Neuroscience*, 2, 1–9.
- Wixted, J. T. (2007). Dual-process theory and signal-detection theory of recognition memory. *Psychological Review*, 114, 152–176.
- Wixted, J. T. (2009). Remember/know judgments in cognitive neuroscience: An illustration of the underrepresented point of view. *Learning & Memory*, 16, 406–412.
- Wixted, J. T., & Stretch, V. (2004). In defense of the signal-detection interpretation of remember/know judgments. *Psychonomic Bulletin & Review*, 11, 616–641.
- Yonelinas, A. P. (1994). Receiver-operating characteristics in recognition memory: Evidence for a dual-process model. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20, 1341–1354.
- Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language*, 46, 441–517.
- Yonelinas, A. P., Otten, L. J., Shaw, K. N., & Rugg, M. D. (2005). Separating the brain regions involved in recollection and familiarity in recognition memory. *Journal of Neuroscience*, 25, 3002–3008.

(Appendix follows)

Appendix

Fitting the CDP Model and the Criterion Variability Model to Empirical Data

To fit a simple Gaussian memory-strength model to confidence-based receiver operating characteristic data (which we briefly describe first to introduce relevant notation), the mean and standard deviation of the lure distribution are arbitrarily set to 0 and 1, respectively. For a 6-point confidence scale, the adjustable parameters include five confidence criteria ( $c_2$  through  $c_6$ , where the subscript represents the confidence rating supplied when memory strength,  $S$ , just exceeds the criterion) and the mean and standard deviation of the target distribution ( $\mu_T$  and  $\sigma_T$ , respectively). Figure A1 depicts this model.

For a target, the predicted probability of observing a confidence rating of  $i$  or greater is

$$P(S > c_i | \text{Target}) = \Phi\left(\frac{\mu_T - c_i}{\sigma_T}\right),$$

where  $2 \leq i \leq 6$ ,  $\mu_T$  and  $\sigma_T$  represent the mean and standard deviation of the target distribution, and  $\Phi$  is the normal cumulative distribution function. For a lure, the predicted probability of observing a confidence rating of  $i$  or greater (where, again,  $2 \leq i \leq 6$ ) is

$$P(S > c_i | \text{Lure}) = \phi(-c_i).$$

For both targets and lures, the predicted probability of observing a confidence rating of 1 or more is, of course, 1.0.

These equations are used to predict observed confidence ratings of 1 through 6 for a specific set of parameter values, from which the log likelihood of the observed data can be computed. An optimization routine is then used to adjust the parameters until the log likelihood of the data given the model is maximized (or, equivalently,  $G^2$  is minimized). The same strategy is used to fit the continuous dual-process (CDP) and criterion-variability models

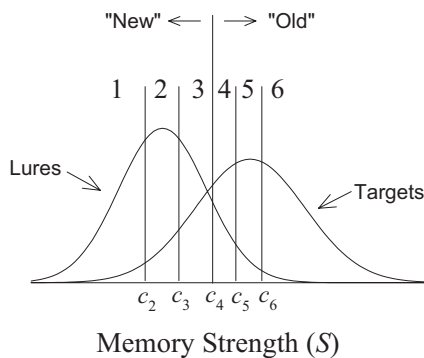


Figure A1. An illustration of the unequal-variance signal-detection model showing the location of five confidence criteria ( $c_2$  through  $c_6$ ).

once the relevant model equations are specified. We describe those equations next.<sup>3</sup>

To fit the CDP model to data reported by Rotello et al. (2005), the goal is to find predicted probabilities for remember and know judgments for different levels of confidence,  $i$ , ranging from 2 through 6 (confidence ratings of 1 were equivalent to “new” decisions and were not accompanied by remember/know judgments). Figure A2 represents the CDP model in two-dimensional space.

The  $x$ -axis in this figure corresponds to recollection ( $R$ ), and the Gaussian distribution placed on the  $x$ -axis represents the distribution of  $R$  for a class of test items (either targets or lures). The  $y$ -axis corresponds to the sum of recollection ( $R$ ) and familiarity ( $F$ ) for the same class of test items. The series of vertically oriented distributions represent the distribution of  $F$  displaced by the corresponding value of  $R$  on the  $x$ -axis (there are an infinite number of these displaced distributions). The mean of a particular vertically oriented distribution (i.e., the distribution of  $F$  displaced by a particular value of  $R$ ) is either  $\mu_{F0} + R$  (for lures) or  $\mu_{F1} + R$  (for targets), and its standard deviation is 1 (because  $\sigma_{F0}$  and  $\sigma_{F1}$  both equal 1). The value of  $c_i$  is a particular confidence criterion on the memory-strength ( $R + F$ ) axis, and the value of  $r$  is the decision criterion for remember judgments on the recollection ( $R$ ) axis. The proportion of remember judgments for a confidence rating of  $i$  or greater,  $P(F + R > c_i \text{ and } R > r)$ , corresponds to the mass in the top right quadrant, and the proportion of know judgments for a confidence rating of  $i$  or greater corresponds to the mass in the top left quadrant. For a target item, the probability that aggregate memory strength ( $F + R$ ) exceeds the decision criterion ( $c_i$ ) is

$$P(F + R > c_i) = \phi\left(\frac{\mu_{F1} + R - c_i}{\sigma_{F1}}\right).$$

To find the proportion of items for which  $F + R > c_i$  and  $R > r$  (and so receive remember judgments), the above quantity is multiplied by the relative likelihood of observing a particular value of  $R$ , and the entire expression is integrated with respect to  $R$  over the range of  $r$  to infinity. For remember judgments to targets, the expression would be

$$P(F + R > c_i \text{ and } R > r | \text{Target}) = \int_r^\infty \Phi\left(\frac{\mu_{F1} + R - c_i}{\sigma_{F1}}\right) \cdot \phi\left(\frac{R - \mu_{R1}}{\sigma_{R1}}\right) dR,$$

<sup>3</sup> We thank John Dunn for deriving the equations for the CDP and criterion-variability models, for explaining the derivations in the manner we describe here, and for writing the MATLAB routines we used to fit those models to the data from Rotello et al. (2005).

(Appendix continues)

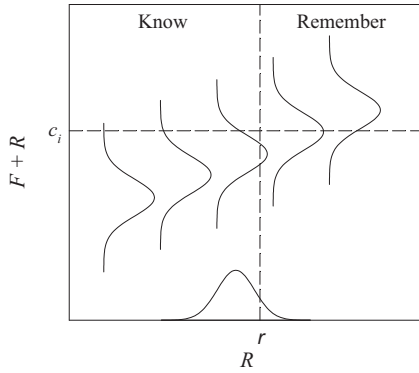


Figure A2. A two-dimensional illustration of the continuous dual-process (CDP) model.

where  $\phi$  is normal probability density function. For remember judgments to lures, the corresponding expression would be

$$P(F + R > c_i \text{ and } R > r | \text{Lure}) = \int_r^\infty \phi\left(\frac{\mu_{F0} + R - c_i}{\sigma_{F0}}\right) \cdot \phi\left(\frac{R - \mu_{R0}}{\sigma_{R0}}\right) dR.$$

Note that  $r, c_2, c_3, c_4, c_5,$  and  $c_6$  represent decision criteria, and all are free parameters. In addition,  $\mu_{F1}, \mu_{R1},$  and  $\sigma_{R1}$  are free parameters, whereas  $\mu_{F0}$  and  $\mu_{R0}$  are both set to 0, and  $\sigma_{R0}, \sigma_{F0},$  and  $\sigma_{F1}$  are all set to 1 (where the subscripts 0 and 1 refer to lures and targets, respectively). Thus, these equations simplify to

$$P(F + R > c_i \text{ and } R > r | \text{Target}) = \int_r^\infty \phi(\mu_{F1} + R - c_i) \cdot \phi\left(\frac{R - \mu_{R1}}{\sigma_{R1}}\right) dR,$$

$$P(F + R > c_i \text{ and } R > r | \text{Lure}) = \int_r^\infty \phi(R - c_i) \cdot \phi(R) dR.$$

The equations for know judgments to targets and lures are similar, except that now we are interested in items for which aggregate memory strength exceeds  $c_i$  (i.e.,  $F + R > c_i$ ) but recollection falls below the recollection criterion (i.e.,  $R < r$ ). Thus, the relevant expressions are integrated from 0 to  $r$  with respect to  $R$ :

$$P(F + R > c_i \text{ and } R < r | \text{Target}) = \int_0^r \phi(\mu_{F1} + R - c_i) \cdot \phi\left(\frac{R - \mu_{R1}}{\sigma_{R1}}\right) dR,$$

$$P(F + R > c_i \text{ and } R < r | \text{Lure}) = \int_0^r \phi(R - c_i) \cdot \phi(R) dR.$$

Again, for all of these cases,  $2 \leq i \leq 6$ . These equations can be used to predict observed frequencies of remember and know judgments for different levels of confidence (and for new decisions associated with confidence ratings of 1) for a given set of parameter values, and an optimization routine can then be used to maximize the likelihood of the observed data given the CDP model.

A very similar set of considerations applied to the criterion-variability model. Figure A3 represents this model. In this case, the  $x$ -axis is unidimensional memory strength ( $S$ ), and the  $y$ -axis is the difference between the remember criterion ( $r$ ) and  $S$ . The sequence of vertically oriented distributions illustrates the distribution of  $r$  displaced by the corresponding value of (minus)  $S$  on the  $x$ -axis. In the absence of criterion variability, the difference between  $r$  and  $S$  would be constant for a particular value of  $S$ . In the presence of criterion variability, however, each value of  $S$  is associated with a distribution of  $r - S$ . The mean of this distribution is  $\mu_r - S$  (where  $\mu_r$  is the mean location of the remember criterion), and the standard deviation of this distribution is  $\sigma_r$  (the standard deviation of the remember criterion). The proportion of remember responses now corresponds to the mass in the lower right quadrant, i.e.,  $S > c$  and  $r - S < 0$ . Thus, the relevant equations for remember judgments to targets and lures are

$$P(S > c_i \text{ and } r - S < 0 | \text{Target}) = \int_{c_i}^\infty \phi\left(\frac{S - \mu_r}{\sigma_r}\right) \cdot \phi\left(\frac{S - \mu_T}{\sigma_T}\right) dS,$$

$$P(r - S < 0 | \text{Lure}) = \int_{c_i}^\infty \phi\left(\frac{S - \mu_r}{\sigma_r}\right) \cdot \phi\left(\frac{S - \mu_L}{\sigma_L}\right) dS.$$

As in the CDP model,  $\mu_r$  (now the mean location of the remember criterion),  $c_2, c_3, c_4, c_5,$  and  $c_6$  represent decision criteria, and they are all free parameters. In addition,  $\mu_T$  (the mean of the target distribution),  $\sigma_T$  (the standard deviation of the target distribution),

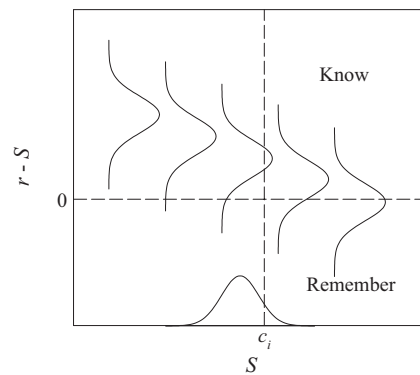


Figure A3. A two-dimensional illustration of the criterion variability model.

(Appendix continues)

and  $\sigma_r$  (the standard deviation of  $r$ ), are free parameters, whereas  $\mu_L$  and  $\sigma_L$  (the mean and standard deviation of the lure distribution, respectively) are set to 0 and 1, respectively. Thus, the equation for lures simplifies to

$$P(S > c_i \text{ and } r - S < 0 | \text{Lure}) = \int_{c_i}^{\infty} \phi\left(\frac{S - \mu_r}{\sigma_r}\right) \cdot \varphi(S) dS.$$

Finally, the relevant equations for know judgments to targets and lures are

$$P(S > c_i \text{ and } r - S > 0 | \text{Target})$$

$$= \int_{c_i}^{\infty} \phi\left(\frac{\mu_r - S}{\sigma_r}\right) \cdot \varphi\left(\frac{S - \mu_T}{\sigma_T}\right) dS,$$

$$P(S > c_i \text{ and } r - S > 0 | \text{Lure}) = \int_{c_i}^{\infty} \phi\left(\frac{\mu_r - S}{\sigma_r}\right) \cdot \varphi(S) dS.$$

Received December 21, 2009

Revision received June 4, 2010

Accepted June 7, 2010 ■

### Call for Nominations

The Publications and Communications (P&C) Board of the American Psychological Association has opened nominations for the editorships of **Journal of Experimental Psychology: Learning, Memory, and Cognition**; **Professional Psychology: Research and Practice**; **Psychology and Aging**; **Psychology, Public Policy, and Law**; and **School Psychology Quarterly** for the years 2013–2018. Randi C. Martin, PhD, Michael C. Roberts, PhD, Ronald Roesch, PhD, and Randy W. Kamphaus, PhD, respectively, are the incumbent editors.

Candidates should be members of APA and should be available to start receiving manuscripts in early 2012 to prepare for issues published in 2013. Please note that the P&C Board encourages participation by members of underrepresented groups in the publication process and would particularly welcome such nominees. Self-nominations are also encouraged.

Search chairs have been appointed as follows:

- **Journal of Experimental Psychology: Learning, Memory, and Cognition**, Leah Light, PhD, and Valerie Reyna, PhD
- **Professional Psychology: Research and Practice**, Bob Frank, PhD, and Lillian Comas-Diaz, PhD
- **Psychology and Aging**, Leah Light, PhD
- **Psychology, Public Policy, and Law**, Peter Ornstein, PhD, and Brad Hesse, PhD
- **School Psychology Quarterly**, Neal Schmitt, PhD, and Jennifer Crocker, PhD

Candidates should be nominated by accessing APA's EditorQuest site on the Web. Using your Web browser, go to <http://editorquest.apa.org>. On the Home menu on the left, find "Guests." Next, click on the link "Submit a Nomination," enter your nominee's information, and click "Submit."

Prepared statements of one page or less in support of a nominee can also be submitted by e-mail to Sarah Wiederkehr, P&C Board Search Liaison, at [swiederkehr@apa.org](mailto:swiederkehr@apa.org).

Deadline for accepting nominations is January 10, 2011, when reviews will begin.