Psychological Review
2007, Vol. 114, No. 1, 152–176

0033-295X/07/$12.00 DOI: 10.1037/0033-295X.114.1.152

Dual-Process Theory and Signal-Detection Theory of Recognition Memory

John T. Wixted
University of California, San Diego

Two influential models of recognition memory, the unequal-variance signal-detection model and a dual-process threshold/detection model, accurately describe the receiver operating characteristic, but only the latter model can provide estimates of recollection and familiarity. Such estimates often accord with those provided by the remember–know procedure, and both methods are now widely used in the neuroscience literature to identify the brain correlates of recollection and familiarity. However, in recent years, a substantial literature has accumulated directly contrasting the signal-detection model against the threshold/detection model, and that literature is almost unanimous in its endorsement of signal-detection theory. A dual-process version of signal-detection theory implies that individual recognition decisions are not process pure, and it suggests new ways to investigate the brain correlates of recognition memory.

Keywords: signal-detection theory, dual-process theory, recollection, familiarity, recognition memory

Two influential and seemingly incompatible views of recognition memory have existed side by side for decades. One view—the dual-process theory of recognition memory—holds that recognition decisions are based on two processes, namely, recollection and familiarity (Atkinson & Juola, 1973, 1974; Hintzman & Curran, 1994; Jacoby, 1991; Jacoby & Dallas, 1981; Mandler, 1980). Recollection is a relatively slow process that consists of retrieving specific details associated with the prior presentation of an item, whereas familiarity is a relatively fast process that allows one to appreciate the fact that the item was previously encountered even though no contextual detail can be retrieved. The other view—signal-detection theory—holds that recognition decisions are based on the strength of a memory signal in relation to a decision criterion. The prototypical version of signal-detection theory involves two equal-variance Gaussian distributions (one representing targets and the other representing lures) and one decision criterion. Any test item that generates a memory strength exceeding the criterion is declared to be old; otherwise it is declared to be new (as illustrated in Figure 1A). Although the equal-variance detection model is often used to illustrate signal-detection theory, much evidence suggests that a quantitatively more accurate version of the theory is an unequal-variance model in which the standard deviation of the target distribution somewhat exceeds that of the lure distribution (Egan, 1958, 1975; Ratcliff, Sheu, & Gronlund, 1992), as illustrated in Figure 1B.

Because the signal-detection model is compatible with a single-process view of recognition memory (e.g., the single process of familiarity could be what underlies the unidimensional memory strength signal), it is often viewed as standing in direct opposition to the dual-process view (e.g., Bodner & Lindsay, 2003; Cary & Reder, 2003; Gardiner, Ramponi, & Richardson-Klavehn, 2002). An early attempt to reconcile these two accounts was proposed by Atkinson and Juola (1973, 1974). According to their two-criterion model, if the familiarity of a test item falls above a high criterion value or below a low criterion value, then a fast, familiarity-based decision is made (old or new, respectively). If the value instead falls between the two criteria, then a search process is initiated, which, if successful, leads to a slower, recollection-based old decision. Thus, in this model, the subject is thought to resort to recollection as a backup process whenever familiarity fails to provide a clear answer.

A more recent attempt to reconcile these two views of recognition memory was proposed by Yonelinas (1994; Yonelinas, Kroll, Dobbins, Lazzara, & Knight, 1998). His model holds that recollection is a high-threshold process (i.e., recollection either occurs or does not occur), whereas familiarity is a continuous variable that is governed by an *equal*-variance detection model (as shown in Figure 1A). If a target item on a recognition test occasions recollection, then a high-confidence old decision is made. No decision criterion or any other consideration based on signal-detection theory is needed to characterize a recollection process like that. If recollection fails, however, then a familiarity-based decision is made. Because familiarity is a continuous variable, and because both targets and lures have some degree of familiarity associated with them, these decisions are thought to be characterized by the signal-detection process illustrated in Figure 1A. Thus, in this model, the participant is thought to resort to familiarity as a backup process whenever recollection fails to occur.

This dual-process signal-detection/high-threshold theory (henceforth referred to as the DPSD model) has had wide influence, not only in the field of psychology but also in the field of cognitive neuroscience, where results from brain lesion studies and neuroimaging studies are often interpreted in terms of its assumptions (e.g., Aggleton et al., 2005; Cipolotti et al., 2006; Fortin, Wright, & Eichenbaum, 2004; Yonelinas et al., 2002). However, an important and seemingly underappreciated consideration is that the acceptance and use of the DPSD model in these studies necessarily entails the rejection of the classic unequal-variance signal-detection (UVSD) model—a model that has survived de-

I thank Larry Squire for his thoughtful comments on an earlier version of this article.

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

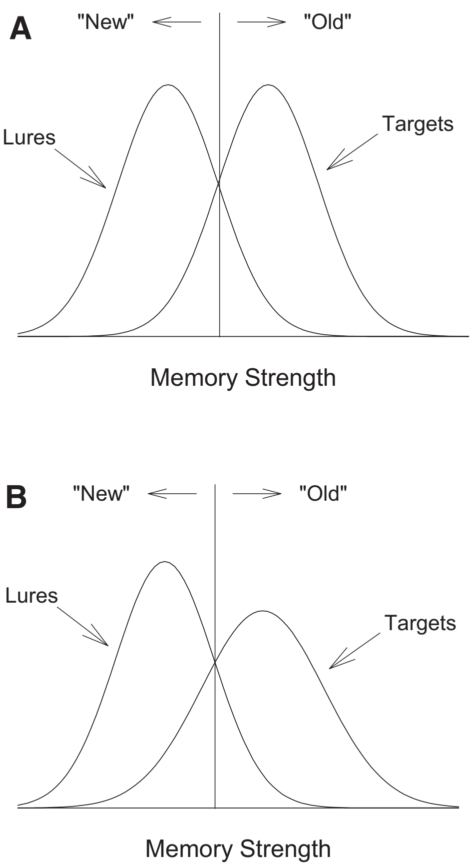

Figure 1. Equal-variance (Figure 1A) and unequal-variance (Figure 1B) signal-detection models of recognition memory.

cades of serious scrutiny. As such, conclusions that are based on the DPSD model are valid only insofar as its underlying assumptions are more viable than those of the UVSD model. But are they? A large body of evidence has accumulated in the past 5 years directly pitting the predictions of the UVSD and DPSD models against each other, and the purpose of this article is to review that literature and to suggest a verdict. As described in detail below, the evidence strongly supports the classic UVSD model. Moreover, it seems clear that the successes of the DPSD model depend almost entirely on the fact that it partially incorporates the assumptions of detection theory (to capture the familiarity process), whereas its inadequacies arise precisely because of the ways in which it departs from detection theory. It is those departures that allow the model to be used to estimate recollection and familiarity. Accordingly, serious reservations must attend estimates of recollection and familiarity when those estimates are derived from the DPSD model.

In what follows, the UVSD and DPSD models are described in more detail (along with an explanation of why they emerged as serious competitors), and then the large and growing body of research that differentially evaluates their predictions is reviewed. Because the signal-detection model is strongly supported by the relevant research, we are ultimately left with the predicament alluded to earlier. Specifically, how can two well-established theories of recognition memory—dual-process theory and signal-detection theory—be reconciled? That question is addressed after the models are comparatively evaluated.

The Competing Models

The UVSD Model

For more than 30 years, signal-detection theory was a dominant theoretical framework for understanding how participants make decisions on recognition memory tasks. It replaced an earlier theory, one that was simple and intuitively compelling, known as *high-threshold theory* (Green & Swets, 1966; Macmillan & Creelman, 2005). High-threshold theory was advanced before recognition memory was thought to involve two processes, but it is described in some detail here because it has been resurrected as part of the DPSD model. The original high-threshold model held that recognition is a probabilistic process. That is, according to this model, a test item is either recognized (i.e., it falls above a threshold) or it is not (i.e., it falls below a threshold), with no degrees of recognition occurring between these extremes. Only target items can generate an above-threshold recognition response because only they appeared on the list. The lures, along with any targets that are forgotten, fall below threshold, which means that they generate no memory signal whatsoever. For these items, the participant has the option of declaring them to be new (as a conservative participant might do) or guessing that some of them are old (as a more liberal participant might do). False alarms in this model reflect memory-free guesses that are made to some of the lures.

This simple and intuitively appealing model yields the once widely used *correction for guessing formula,* and it predicts a linear receiver operating characteristic (ROC), as described by Green and Swets (1966). An ROC is simply a plot of the hit rate versus the false alarm rate for different levels of bias. A typical ROC is obtained by asking participants to supply confidence ratings for their recognition memory decisions. Several pairs of hit and false alarm rates can then be computed by cumulating ratings from different points on the confidence scale (beginning with the most confident responses). The high-threshold model of recognition memory predicts that a plot of the hit rate versus the false alarm rate (i.e., the ROC) will be linear, as illustrated in Figure 2A.

Signal-detection theory assumed a preeminent position in the field of recognition memory in large part because its predictions about the shape of the ROC were almost always shown to be more accurate than the predictions of the intuitively plausible high-threshold model. More specifically, the signal-detection model, which assumes that memory strength is a graded phenomenon (not a discrete, probabilistic phenomenon) predicts that the ROC will be curvilinear, and because every recognition memory ROC analyzed between 1958 and 1997 was curvilinear, the high-threshold model was abandoned in favor of signal-detection theory. A curvilinear ROC that is consistent with the predictions of signal-detection theory is illustrated in Figure 2C. Although signal-detection theory predicts a curvilinear ROC when the hit rate is plotted against the false alarm rate, it predicts a linear ROC when the hit and false alarm rates are converted to z scores (yielding a z-ROC). The high-threshold model, by contrast, predicts that the z-ROC will be curvilinear. Figures 2B and 2D show the z-ROCs that correspond to the probability ROCs in Figures 2A and 2C, respectively.

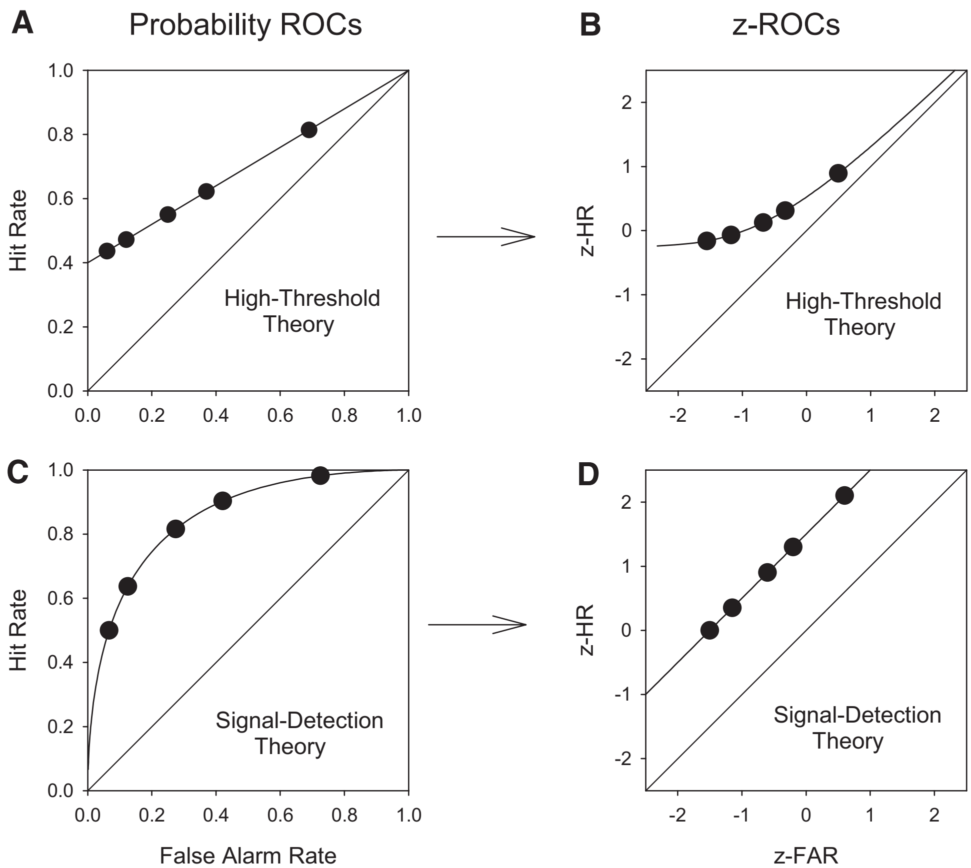

Figure 2. Idealized receiver operating characteristic (ROC) data predicted by the high-threshold model (Figure 2A) and by the signal-detection model (Figure 2C). Figures 2B and 2D show the corresponding ROCs in *z*-space. HR = hit rate; FAR = false alarm rate.

According to the signal-detection model, the slope of the *z*-ROC line provides an estimate of the ratio of the standard deviation of the lure distribution to the standard deviation of the target distribution ($\sigma_{lure}/\sigma_{target}$). If an equal-variance model applies (as in Figure 1A), then the slope should be 1.0, as it is in Figure 2D. But if the standard deviation of the target distribution exceeds that of the lure distribution (as Figure 1B), then the slope of the *z*-ROC should be less than 1.0. Previous meta-analyses of confidence-based ROC data generally show that *z*-ROCs are well characterized by a straight line and that the slope of the best fitting line is less than 1.0 (Glanzer, Kim, Hilford, & Adams, 1999; Ratcliff, Sheu, & Gronlund, 1992). The slope often decreases as accuracy increases (Glanzer et al., 1999), but a common value for the slope is approximately 0.80 (Ratcliff et al., 1992). Thus, according to the signal-detection account, the standard deviation of the target distribution is typically about 1.25 (i.e., 1/0.80) times that of the lure distribution. Findings like these explain why the UVSD model shown in Figure 1B (not the equal-variance model shown in Figure 1A) is regarded by many as the standard model of decision making on a recognition memory task.

It might seem that the UVSD model is inherently less plausible than the more aesthetically appealing equal-variance model, but the opposite is actually true. The targets can be thought of as lures that have had memory strength added to them by virtue of their appearance on the study list. An equal-variance model would result if each item on the list had the exact same amount of strength added during study. However, if the amount of strength that is added differs across items, which surely must be the case, then both strength and variability would be added, and an unequal-variance model would apply. Thus, it is actually the equal-variance model that is, a priori, the less plausible account. The empirical ROC data reinforce the idea that items on the list have varying amounts of strength added to them during study.

The DPSD Model

Although the UVSD model offered a viable account of curvilinear ROCs and linear *z*-ROCs, the notion that decisions are based on a unidimensional memory strength variable was not easily reconciled with the dual-process theory of recognition memory. Various dual-process theories, on the other hand, were never specified in enough quantitative detail to allow them to make clear predictions about the shape of the ROC. A model proposed by Yonelinas (1994) was the first dual-process model that made clear predictions about (and provided a good description of) curvilinear ROCs. This model incorporated high-threshold theory to explain one process (recollection) and incorporated the equal-variance signal-detection model to explain the other. According to this account, item recognition is based on recollection whenever recollective strength exceeds a *threshold*; otherwise it is based on familiarity. The familiarity process is characterized by an equal-variance signal-detection model, but the frequent occurrence of recollection changes the shape of what would otherwise be a symmetrical curvilinear ROC with a z-ROC slope of 1.0. Thus, the fact that the slope of the z-ROC for item recognition is typically less than 1.0 and the fact that the ROC is asymmetrical are explained not by assuming an UVSD model but instead by assuming that some responses are based on threshold recollection, which yields a high-confidence old decision, whereas others are based on an equal-variance detection process, which yields old decisions with varying degrees of confidence. Yonelinas (1999b) suggested that this model fits ROC data as well as the standard UVSD model. A noteworthy advantage of the DPSD model is that it yields quantitative estimates of recollection and familiarity when it is fit to ROC data, and Yonelinas (2002) argued that those estimates approximately correspond to estimates of recollection and familiarity that are provided by the remember–know procedure.

Both the UVSD and the DPSD models provide a good description of curvilinear probability ROCs, and to the naked eye, those fits usually appear to be equally good. However, their predictions about the shape of the ROC are not identical, and the differences between them are easier to see in the z-ROC. As indicated earlier, signal-detection theory predicts a linear z-ROC, but the DPSD model predicts that the z-ROC will be at least slightly curvilinear (i.e., it will be slightly U-shaped) whenever recollection plays a role. Yonelinas (1999b) argued that z-ROCs do, indeed, exhibit the curvilinearity predicted by his model. That curvilinearity is not as exaggerated as the curvilinearity predicted by the standard high-threshold theory (Figure 2B), because in the DPSD model, the threshold recollection process underlies only some of the recognition decisions, whereas familiarity underlies others.

An especially decisive test of this model would be to find a recognition task that involved only one of the two processes. A recognition task that involved only the recollection process, for example, should yield a linear ROC (and an especially curvilinear z-ROC), whereas a task that involved only the familiarity process should yield a symmetrical curvilinear ROC (and a linear z-ROC with a slope of 1.0). A candidate for a recollection-only recognition task is the associative recognition procedure in which participants are first presented with a list of word pairs and then asked to discriminate intact word pairs from rearranged word pairs on the recognition test (e.g., Hockley, 1991). Item familiarity offers no help on this task because all of the items are familiar, so it is ordinarily assumed that the only way to solve the task is to recollect the word's associate at study. If recollection is a discrete, high-threshold process, then the ROC should be linear and the z-ROC should be U-shaped. Yonelinas (1997) reported precisely that result. This was the first linear ROC ever reported for any recognition memory task, and it provided compelling evidence that recognition decisions are sometimes based on a threshold recollection process. Without some modification, the standard UVSD model never predicts a linear ROC.

A similar story played out in the source-monitoring literature. In a typical source-monitoring task, participants hear some words from one source and other words from another source (e.g., Johnson, Hashtroudi, & Lindsay, 1993). For example, the words might be presented by two different speakers, one male and one female. On a later recognition test, the words are presented visually, and participants are asked to decide whether the words are old or new (a standard recognition task that theoretically involves both recollection and familiarity). For each word that is declared to be old, they are further asked to indicate its source (male or female speaker). Familiarity offers no help on the source task because the words from both sources are, on average, equally familiar, so relying on recollection is generally assumed to be the only way to solve the task. If recollection is a high-threshold process, then the source recognition ROC should also be linear. Yonelinas (1999a) reported findings that confirmed this prediction.

The studies discussed here theoretically isolated the recollection process and found that the ROC was linear, as uniquely predicted by the DPSD model. Other studies have attempted to isolate the familiarity process by eliminating recollection. According to one theory, the hippocampus selectively subserves the recollection process (Aggleton & Shaw, 1996; Brown & Aggleton, 2001). Thus, individuals with lesions limited to the hippocampus (or rats with hippocampal lesions) should be able to rely only on the familiarity process. In this model, familiarity is governed by the equal-variance signal-detection model shown in Figure 1A. As such, the performance of individuals with hippocampal lesions should be characterized by a *symmetrical* curvilinear ROC (and a linear z-ROC with a slope of 1.0). A few studies of humans or rats have reported evidence consistent with this prediction (Aggleton et al., 2005; Fortin et al., 2004; Yonelinas et al., 1998, 2002).

Comparative Evaluation

The competition between the UVSD model and the DPSD model is an interesting one because they are both characterized by two parameters, and they both fit ROC data well. Even so, their assumptions are quite different. Of most importance in this regard, the DPSD model, like all dual-process models before it, assumes that a given recognition decision is process pure, which is to say that the decision is based either on recollection or on familiarity (never on both processes together). The process-purity assumption typically refers to the notion that a memory task in which many items are tested involves only one process (Jacoby, 1991). Here, I instead consider the process-purity assumption as it applies to a decision about an individual test item. Thus, for example, if the test item *diamond* is recognized as being old, the assumption made by all dual-process models (including the DPSD model) is that the participant either recollected the item in sufficient detail to warrant a high-confidence old decision (in which case its degree of familiarity was irrelevant) or did not recollect the item and resorted to the familiarity process instead (in which case recollection played no role). The either/or character of the DPSD model derives from its assumption that recollection occurs in an essentially all-or-none threshold fashion. As Yonelinas (1994) put it,

> In recognition memory, the basic idea is that judgments can be based upon an assessment of item familiarity or on the product of a conscious recollection process. Recollection is assumed to be an all-or-none retrieval process, such that for any item the subject either succeeds or fails at retrieving something about that specific study event. A successful retrieval is expected to lead to a highly confident response. Familiarity, on the other hand, is assumed to be well described by the standard equal-variance signal detection theory described earlier. (p. 1343)

Yonelinas, Dobbins, Szymanski, Dhaliwal, and King (1996) pointed out that the all-or-none characterization of the recollection process can be misleading because above-threshold recollection might be associated with different degrees of high strength. However, according to the DPSD model, the behavioral effects of recollection are as they would be if the process were all or none. For example, when the DPSD model is fit to ROC data to extract estimates of recollection and familiarity, recollection is assumed to yield only high-confidence old decisions (just as an all-or-none model would assume), whereas only familiarity plays a role in decisions made with lower degrees of confidence. Thus, recollection in the DPSD model can be thought of as an all-or-none process in the sense it leads to a high-confidence old decision (not varying degrees of confidence).

In contrast to the DPSD model, the UVSD model holds that each recognition decision is based on a continuously distributed memory strength variable. When memory strength exceeds a criterion value, the item is declared to be old; otherwise it is declared to be new. However, it should not be assumed that the single, unidimensional memory strength variable in the UVSD model necessarily implies a single underlying memory process. Kelley and Wixted (2001), Wixted and Stretch (2004), and Rotello, Macmillan, and Reeder (2004) argued that the memory strength variable in question consists of the additive combination of familiarity and recollection (or, equivalently, item and associative information). That is, when making a decision about an individual test item, the participant is assumed to take both sources of information into account, much like a juror who combines several sources of evidence into an overall assessment of the defendant's guilt (instead of relying on either one source of evidence or the other). This account differs from all other dual-process theories in that it assumes that memory itself is not process pure. Instead, this account assumes that both processes play a role in specific decisions about particular test items even when those decisions are made with lower levels of confidence. For the moment, these points are being raised only to make it clear that the following comparison between the DPSD and UVSD models is not a comparison between a dual-process model and a single-process model.

Old–New ROCs

Which model fits better? As described earlier, both the UVSD model and the DPSD model do a good job of fitting asymmetrical curvilinear ROCs derived from a standard old–new recognition task. Although they both fit well, it is useful to ask which model fits better. This is an important question to consider because it is unlikely that the model with the more accurate assumptions will consistently provide the worse fit. Instead, a model that almost always provides a worse fit, even if it fits reasonably well, usually does so because its assumptions are wrong.

This basic point is worth emphasizing because the ability of these two models to describe old–new recognition memory ROCs—the crucible that once dictated the fate of high-threshold theory—has been well documented in recent years, and most results show that the UVSD model far outperforms the DPSD model. This issue was first raised by Yonelinas (1999b) in response to criticisms of the DPSD model advanced by Glanzer et al. (1999). In a section entitled *Directly Testing the Dual-Process and Unequal-Variance Models*, Yonelinas (1999b) stated that he "simply fit the two models to the observed ROCs" (p. 517) and that "on average, the dual-process model accounted for 99.91% of the variance, and the unequal-variance model accounted for 99.97%" (p. 517). He went on to note that both models fit so well that this approach may not be a fruitful way to discriminate between them. Although that may have been true of the few group ROCs that were considered in that study, it has not proven to be true of other analyses, most of which document the clear superiority of the UVSD account.

Kelley and Wixted (2001) found that the DPSD model exhibited significant deviations from the item recognition ROC in two of four conditions they investigated, whereas the UVSD fit well in all four cases (i.e., the UVSD model provided the better fit). However, these analyses were performed on group ROCs, and a comparison of two models that yield similar fits is more convincing when performed at the level of the individual participant to ensure that the results are not influenced by averaging artifacts. Heathcote (2003) recently performed a comprehensive analysis of the ability of these two models to fit individual-participant ROC data. Across three experiments, he found that the UVSD model provided a better fit than the DPSD model for 75% to 80% of the individual ROCs. Heathcote noted that although both models fit the data well, as Yonelinas (1999b) previously found, a good fit in and of itself is not a compelling argument in favor of a theory (Roberts & Pashler, 2000). This point is especially important in this case because confidence-based ROC data represent nonindependent data points. That is, moving from the bottom left to the upper right of the ROC, each new point includes all of the observations that comprised the previous point (plus some additional observations). This lack of independence greatly reduces variability in the data, which makes it much easier for a two-parameter curvilinear function to account for a high percentage of the data variance. Thus, to distinguish between valid and invalid models, one must consider which model reliably provides the better fit instead of focusing on the amount of data variance each model can explain. As Heathcote showed, the UVSD model provides the better fit in the large majority of cases.

Healy, Light, and Chung (2005) also compared the fits of both models at the level of the individual-participant ROC and reported the same result. That is, the UVSD model provided a better fit to item recognition ROCs than the DPSD model by better than two to one, and this was found to be true for both younger and older participants.

Rotello, Macmillan, Hicks, and Hautus (in press) provided yet another analysis along these lines, one that included fits of an interesting modification of the DPSD model. In one representative fit to the individual ROC data, the UVSD model provided the best fit for 19 of 22 participants, confirming the previous findings of Heathcote (2003) and Healy et al. (2005). Rotello et al. also tested a different version of the DPSD model, which was made possible because their participants supplied remember–know judgments as well as confidence ratings. In the extended version of the DPSD model, they allowed recollection responses to be distributed over all levels of confidence that an item was old (whereas ordinarily the model assumes that recollection leads to a high-confidence old decision). Recollection was still assumed to preempt any role for familiarity (i.e., individual recognition decisions were still regarded as being process pure), but recollection could be associated with lower levels of confidence. This modification improved the fit of the DPSD model, but it still did not fit anywhere near as well as the UVSD model in that only a few participants produced data that were best fit by the extended model. As Rotello et al. observed,

> Of course, the advantage conferred by the extended model is to soften the high-threshold nature of the recollection process by allowing remember responses to be distributed across ratings. In this way, the improvement in fit occurs because the extended dual-process model is more like the one-dimensional model.

That is, the more similar to the UVSD model the DPSD model becomes, the better able it is to fit ROC data.

Howard, Bessette-Symons, Zhang, and Hoyer (2006) recently reported the first direct comparison between the ability of the two models to describe old–new ROC data that favored the DPSD model over the UVSD model. They presented a list of pictures depicting intricate travel scenes to both young and old participants and found that the DPSD model fit the individual-participant data reliably better than the UVSD model. Howard et al. speculated that it was the use of those pictures instead of words that accounted for the difference between their results and the rest of the literature, although it is not clear why the choice of stimulus would matter. The pictures were presented at a rapid rate (one per 1.5 s), and it seems possible that participants might not have encoded some of the complex pictures under those conditions. If so, and as described in more detail later, ROC anomalies favoring the DPSD model might be expected (DeCarlo, 2002; Malmberg & Xu, 2006). In any case, on the whole, the literature shows that in numerous studies involving words as stimuli, ROC analyses clearly favor the UVSD model (despite initial claims to the contrary). For rapidly presented complex travel scenes, one recent study supports the DPSD model.

Model flexibility. The fact that the DPSD model is, in most cases, not competitive with the UVSD model in terms of its ability to fit ROC data should decrease confidence in its validity. However, there is one scenario under which the model that provides the consistently poorer fit is nevertheless the more valid model. Specifically, even if models are equated for the number of free parameters, as these two models are, they are not necessarily equated in terms of flexibility (Pitt, Kim, & Myung, 2003). Flexibility refers to a model's ability to adjust itself to outcomes that arise from processes that are not embraced by the model. A highly flexible model would be underconstrained and would, for example, be able to adjust itself to random error in the data or to fit simulated data that were generated by an altogether different model. Thus, high flexibility is an undesirable property, and it is possible that the reason the UVSD model so reliably outperforms the DPSD model is simply because it is the more flexible of the two models, not because it is the more valid of the two. One way to gain insight into the relative flexibility of these models is to generate simulated data from each model separately and then to fit both models to those simulated data (i.e., to perform a model recovery simulation, as described by Pitt et al., 2003). Ideally, the UVSD model would tend to provide the best fit when the data are generated by that model, and the DPSD model would tend to provide the best fit when the data are generated by that model. On the other hand, if the UVSD model tended to provide the best fit no matter which model actually produced the data, then its clear superiority in fitting individual ROC data would be hard to interpret.

I investigated this issue by generating simulated data from both models, first under conditions in which overall memory performance was relatively low and then again when it was somewhat higher. The signal-detection simulations involved randomly drawing 200 memory strength observations from the target distribution and 200 from the lure distribution. In the low-strength condition, the target distribution had a mean of 1.5 and a standard deviation of 1.25, whereas the lure distribution had a mean of 0 and a standard deviation of 1.0. Responses drawn from these distributions were classified into confidence bins on the basis of their relationship to five confidence criteria (the confidence criteria were placed 1.6, 1.2, 0.8, 0.3, and −0.1 standard deviations from the mean of the lure distribution). The 400 observations drawn from the target and lure distributions constituted 1 simulated ROC, and the process was repeated until 30 simulated ROCs were produced. Each ROC was then fit by both the UVSD model (which is the model that generated the data) and the DPSD model. Next, 30 ROCs were generated from the DPSD model by setting r (the recollection parameter) to a probability of .17 and d' (the familiarity parameter) to 1.0. These settings yielded ROC data that were very much like the ROC data produced by the signal-detection simulation. Again, both models were fit to these 30 ROCs using a standard maximum likelihood estimation procedure.

The integer values in Table 1 show how often each model fit best (i.e., yielded the lower chi-square value) as a function of which model produced the data. As shown in the table, the UVSD model provided the best fit about two thirds of the time when the data were generated by that model, and the DPSD model provided the best fit about two thirds of the time when the data were generated by that model. That is, usually the model that generated the data provided the best fit, which is as it should be. Table 1 also shows the chi-square goodness-of-fit results from these fits (a lower chi-square indicates a better fit). Because there are 3 degrees of freedom associated with each fit, the expected chi-square is 90 summed over the 30 fits. The correct model produced the lower overall chi-square value each time (which, again, is as it should be).

A similar test was performed on simulated data from a stronger memory condition. For the UVSD simulations, the target distribu-

Table 1
Number of Times (Out of 30) the DPSD and UVSD Models Best Fit Simulated Weak Memory Data Generated by a DPSP and a UVSD Model

	Fitted model	
True model	DPSD	UVSD
DPSD		
No.	20	10
Chi-square	$\chi^2(90) = 98.5$	$\chi^2(90) = 126.2^*$
UVSD		
No.	10	20
Chi-square	$\chi^2(90) = 87.4$	$\chi^2(90) = 77.9$

Note. The chi-square values show the goodness-of-fit statistic summed over each of the 30 fits. $N = 400$ for all chi-square values. DPSD = dual-process signal-detection/high-threshold model; UVSD = unequal-variance signal-detection model.
$^* p < .05$.

tion had a mean of 2.0 and a standard deviation of 1.25. For the DPSD simulations, the parameters were set to $d' = 1.4$ and $r = .35$. Table 2 show that, once again, the model that generated the data is the one that fit best approximately two thirds of the time. Note that the model that generated the data does not fit the best 100% of the time because of random error in the data. Had more than 400 observations been used, the correct model would have fit the data best more than two thirds of the time.

While this analysis is not exhaustive, it does suggest that in the range of simulated data (which is in a range that is fairly typical of real data), there is no evident bias to recover one model over the other. That is, there is no apparent bias favoring the detection model even when the data were actually produced by the DPSD model (or vice versa). Thus, the results suggest that when fitting real data, the detection model offers the best fit most of the time because it is more viable than the DPSD model. It should be emphasized that when the UVSD model generated the simulated data (i.e., when it was definitely the case that the DPSD model was the wrong model), the DPSD model fit ROC data well. This finding helps to drive home the point that a model can fit data well even when it is the wrong model. As such, the fact that the DPSD model provides a good fit to empirical ROC data is not, in and of itself, very informative. Although the DPSD model fits well even when the UVSD model generated the data, it does not fit as well as the UVSD model in the majority of cases. That is precisely the pattern that is usually observed when the two models are fit to real ROC data.

All of this simply confirms what Glanzer et al. (1999) argued in a different way. As indicated above, Glanzer et al. pointed out that with regard to the DPSD model's account of the *z*-ROC, the factor that accounts for the slope being less than 1.0 and the factor that accounts for curvilinearity of the *z*-ROC are one and the same, namely, all-or-none recollection. Thus, whenever the slope is less than 1.0 (theoretically indicating the presence of recollection), the *z*-ROC should be curvilinear. Glanzer et al. directly tested this prediction and showed (a) that this relationship does not hold and (b) that old–new *z*-ROCs do not exhibit any consistent curvilinearity at all. Heathcote (2003) also noted the absence of curvilinear *z*-ROCs in the individual-participant data he analyzed across multiple experiments, although Howard et al. (2006) recently found curvilinear *z*-ROCs for rapidly presented, complex travel scenes. The usual absence of that phenomenon (beyond what would be expected because of random error in the data) is simply another way of making the point that UVSD theory usually provides a better description of old–new ROC data than the DPSD model does.

Table 2
Number of Times (Out of 30) the DPSD and UVSD Models Best Fit Simulated Strong Memory Data Generated by a DPSD and a UVSD Model

True model	Fitted model: DPSD	Fitted model: UVSD
DPSD		
No.	18	12
Chi-square	$\chi^2(90) = 70.2$	$\chi^2(90) = 82.4$
UVSD		
No.	10	20
Chi-square	$\chi^2(90) = 87.4$	$\chi^2(90) = 77.9$

Note. The chi-square values show the goodness-of-fit statistic summed over each of the 30 fits, $N = 400$ for all chi-square values. DPSD = dual-process signal-detection/high-threshold model; UVSD = unequal-variance signal-detection model.

Predicting Forced-Choice Performance From Old–New Performance

The earlier discussion focused on the ability of the two models to, essentially, retrodict the data (i.e., to fit the ROC data after the fact). The results were compelling, but an even more compelling test would be provided by testing the ability of each model to predict a quantitative outcome in advance. Such a test would also help to address the issue of model flexibility because a model that is more flexible (thereby fitting data well) pays a price in its ability to make accurate predictions (Pitt et al., 2003). Smith and Duncan (2004) recently conducted an interesting test along these lines. It is well known that an equal-variance signal-detection model predicts that the relationship between old–new recognition performance ($d'_{\text{O-N}}$) and forced-choice recognition performance (d'_{FC}) will be given by the following equation:

$$d'_{\text{FC}} = \sqrt{2}(d'_{\text{O-N}}). \tag{1}$$

That is, forced-choice (FC) performance should be better than old–new (O-N) performance and by a specified amount ($\sqrt{2}$). The predicted relationship changes only slightly for the UVSD model:

$$d'_{\text{FC}} = 2(d'_{\text{O-N}})/(s^2+1)^{1/2}, \tag{2}$$

where s is the slope of the ROC (a value that is often about 0.80). To test this predicted relationship, one can present participants with a study list and then test half the items using an old–new recognition test and the other half using a forced-choice recognition test. An estimate of s can be obtained from the old–new ROC, and Equation 2 can then be used to derive predicted d' values for the forced-choice test (d'_{FC}). The question of interest is how well the predicted d'_{FC} values correspond to the obtained d'_{FC} values. Smith and Duncan performed just this test of the UVSD model.

Smith and Duncan (2004) also conducted a similar predictive test, using the DPSD model. That is, estimates of recollection and familiarity were first derived by fitting the DPSD model to the old–new ROC, and these values were used to predict estimates of recollection and familiarity derived from fitting the model to the forced-choice recognition performance. The recollection estimate obtained from the old–new task should match the estimate obtained from the forced-choice task (i.e., the probability of recollection should not be affected by the nature of the test), and the familiarity d' estimates should be related by Equation 1 above (not Equation 2, because in this model the familiarity process is governed by an equal-variance detection model). They performed this test in part because an earlier study comparing old–new and forced-choice performance by Kroll, Yonelinas, Dobbins, and Frederick (2002) claimed to find that the relationship between the two tests was better explained by the DPSD model than by UVSD theory. However, Kroll et al. focused most of their criticism on the equal-variance version of signal-detection theory (which has few advocates). In addition, they showed that proportion correct on a forced-choice test could be predicted from fits of the DPSD model to old–new data. However, this is a weak test because many different estimates of recollection and familiarity obtained from fitting the DPSD model to old–new data can predict the same level of performance on a forced-choice task. The more relevant question is whether the estimates of recollection and familiarity derived from the old–new task correspond to the estimates obtained from the forced-choice task. That is, for both the UVSD model and the DPSD model, the question was whether the theoretically significant parameters obtained from one task apply to the other (as they should if the model is valid).

Smith and Duncan (2004) found that d' estimates derived from fitting the UVSD model to old–new item recognition ROCs reliably predicted d' estimates on forced-choice recognition tasks. The variance accounted for across participants was an impressive 66%, a result that provides compelling support for the UVSD model. By contrast, the DPSD model performed much worse, especially the non-signal-detection component (i.e., the threshold recollection component) of that model. More specifically, the predicted familiarity d'_{FC} values captured a respectable 31% of the variance of the obtained familiarity d'_{FC} values across participants. However, the predicted recollection (R_{FC}) values accounted for less than 1% of the variance in the obtained R_{FC} values. As Smith and Duncan observed,

> Of special importance was the finding that an individual subject's retrieval via the recollection process in Y/N [yes/no] is uncorrelated with his or her subsequent retrieval by recollection in 2AFC [two-alternative forced choice]. According to Figure 4C, a group of subjects with observed recollection probabilities near .10 in 2AFC could have varied from near 0 to as high as .60 in their observed Y/N recollection probability. (pp. 622–623)

On the basis of these results, Smith and Duncan concluded that the ability of the DPSD model to fit item recognition ROCs reasonably well derives mainly from its inherent flexibility (not its theoretical validity). This result accords with the relative abilities of the two models to fit individual old–new ROC data, and taken together, the retrodictive and predictive tests suggest that the DPSD model may not be valid.

These results can be considered from another point of view by regarding a symmetrical curvilinear ROC (with a z-ROC slope of 1.0) that corresponds to the equal-variance signal-detection model as a starting point. Then one can consider what happens when different approaches are used to account for the fact that ROCs are typically asymmetrical (with a z-ROC slope of less than 1.0). To do that, one can allow the target distribution to have greater variance than the lure distribution, as the UVSD model does, or one can add an (essentially) all-or-none recollection process, as the DPSD model does. Both approaches allow the models to fit asymmetrical curvilinear ROC data reasonably well. However, the former strategy yields excellent retrodictive and predictive validity, whereas the latter leads to substantially poorer retrodictive validity and extremely poor predictive validity (at least with regard to estimates of recollection). This outcome accords with what Rotello et al. (in press) concluded: The failings of the DPSD model stem primarily from the way in which it construes the recollection process. The idea that recollection always leads to a high-confidence old decision allows the model to be used to extract quantitative estimates of recollection and familiarity from an ROC. However, the assumptions of the DPSD model that allow it to do that are the very assumptions that account for its failings in relation to the UVSD model. This is a key consideration that is addressed next by considering the ability of the two models to fit recollection-based ROC data.

Associative Recognition and Source Memory ROCs

Are they linear? The results discussed earlier were concerned with the ability of the UVSD and DPSD models to fit ROC data from a standard old–new recognition task. But the most compelling findings that were originally advanced in favor of the DPSD model were based on other kinds of recognition tasks. Specifically, the linear associative recognition ROCs reported by Yonelinas (1997) and replicated by Rotello, Macmillan, and Van Tassel (2000), along with the linear source memory ROCs reported by Yonelinas (1999a), appeared to legitimize the DPSD model more than any other findings to date. Even the most ardent proponent of the UVSD model is forced to concede that such findings are persuasive. As Kroll et al. (2002) put it,

> Second, in experiments designed to discriminate between the two models, it is clear that the dual-process model provides a better account of the data than does the unequal-variance model. For example, the dual-process model predicts that under conditions in which performance is expected to rely primarily on recollection, the ROCs should become more linear because recollection is assumed to reflect a threshold process. Linear ROCs have been observed in tests of associative recognition and source recognition (e.g., Rotello, Macmillan, & Van Tassel, 2000; Yonelinas, 1997, 1999; Yonelinas, Kroll, Dobbins, & Soltani, 1999; but see Qin, Raye, Johnson, & Mitchell, 2001), indicating that the unequal-variance model is inconsistent with the recognition data. (p. 252)

However, a great deal of work has accumulated since these linear ROCs were first reported, and it is now quite clear that neither associative recognition nor source memory ROCs are typically linear. Instead, they are typically curvilinear, as the detection account would predict.

Kelley and Wixted (2001) were the first to suggest that ROCs produced by a standard associative recognition task (involving intact vs. rearranged word pairs) may, under fairly typical conditions, be very well described by a UVSD model after all. They simply manipulated the strength of word pairs by presenting some pairs once and other pairs multiple times. The weak pairs yielded an atypical curvilinear ROC (one that was essentially halfway between the linear function predicted by high-threshold theory and the curvilinear function predicted by UVSD theory), but the strong pairs yielded an undeniably curvilinear ROC, one that was very accurately described by the UVSD model. This finding creates a dilemma for the DPSD model because strengthening word pairs should not allow item familiarity to suddenly play a role. No matter how familiar the items are, the items of the rearranged pairs are just as familiar as the items of the intact pairs. Thus, the ROC should reflect better performance, but its shape should remain linear.

Since Kelley and Wixted (2001) first reported their results, many curvilinear associative recognition ROCs have been reported in the literature. Verde and Rotello (2004), for example, reported three associative recognition ROCs, and all three were clearly curvilinear. These data are noteworthy because they are from the only laboratory that replicated the linear ROC result. In addition, Healy et al. (2005) stated that they reviewed 13 associative recognition studies and found that models predicting a curvilinear ROC provided a much better fit than models predicting a linear ROC. They also reported six additional associative recognition ROCs, and all six are better described by a curvilinear UVSD model than a linear threshold model. Thus, despite initial indications to the contrary, the overwhelming weight of evidence indicates that associative recognition ROCs are not linear and are much better fit by a UVSD model than by a linear high-threshold model.

The DPSD model can accommodate a curvilinear associative recognition ROC if one assumes that pairs become unitized as a function of the strengthening manipulation, in which case performance might be based on *pair familiarity* instead of item familiarity (Yonelinas, 1997). This is not a natural prediction of the DPSD model, and further research is needed to determine its viability, but it is one way for the DPSD model to deal with curvilinear associative recognition ROCs. Yonelinas, Kroll, Dobbins, and Soltani (1999), for example, appealed to this explanation to account for curvilinear associative recognition ROCs that were observed when pairs of faces were used instead of word pairs. If one allows for this version of the DPSD model, a natural question to ask is whether it fits curvilinear associative recognition ROCs better than the UVSD model. Healy et al. (2005) compared their fits and found that the UVSD model or a some-or-none variant of that model (Kelley & Wixted, 2001) fit their associative recognition ROCs noticeably better than several versions of the DPSD model that allow for curvilinear ROCs. Thus, not only are associative recognition ROCs not linear, they are not as well described by versions of the DPSD model that allow for curvilinearity (because of pair familiarity) as by the UVSD model. This is the same result that has usually been observed for old–new recognition ROCs.

A similar and even more compelling story has played out in the source recognition literature. When Yonelinas (1999a) reported that source recognition tasks yield a linear ROC, the idea that a threshold recollection process plays an important role in recognition memory seemed to attain secure footing. However, Slotnick, Klein, Dodson, and Shimamura (2000) reported a number of source memory ROCs that were clearly curvilinear and that were clearly better fit by the UVSD model than by a threshold model. Qin, Raye, Johnson, and Mitchell (2001) also quickly challenged the idea that the source recognition ROC is linear. Like Slotnick et al., they reported that their source-monitoring experiments yielded undeniably curvilinear ROCs. Moreover, the *z*-ROCs were linear and exhibited no hint of the U-shaped function that ought to exist according to the DPSD model whenever recollection plays a role (and especially when recollection is exclusively involved, as it theoretically is in the case of source monitoring). Qin et al. used much richer source contexts than is typically the case and concluded that “source ROCs are typically curvilinear, except perhaps when the source information available for most items is very impoverished” (p. 1114). Their conclusion left the door open to the possibility that some source recognition ROCs are linear. However, even this possibility has been questioned by Hilford, Glanzer, Kim, and DeCarlo (2002). They reported a series of experiments involving typically impoverished source memory tasks (e.g., “Was the item presented at the top of the screen or bottom of the screen?”) and found that source recognition ROCs were virtually always curvilinear. In fact, in 13 of 14 studies they reviewed, the ROCs clearly exhibited curvilinearity. Another 8 source recognition ROCs were recently reported by Glanzer, Hilford, and Kim (2004), and all 8 were curvilinear. Thus, once again, the most compelling evidence that source recognition involves threshold recollection (namely, a linear ROC) has not withstood subsequent empirical scrutiny.

ROC anomalies. Associative recognition ROCs do not appear to be linear, as Yonelinas (1997) suggested they were, but they are not always as curvilinear as predicted by the UVSD model. In the experiment reported by Kelley and Wixted (2001), the strong pairs yielded an ROC that was well described only by the UVSD model, but the weak pairs yielded an atypical curvilinear ROC that was about halfway between the linear function predicted by high-threshold theory and the curvilinear function predicted by detection theory. What accounts for the odd shape of the ROC in the weak condition?

The very same question arises with regard to source memory tasks in which participants are first asked to decide if the test item is old or new and then asked to state whether any item that was declared to be old was from Source A (e.g., spoken in a female voice) or Source B (e.g., spoken in a male voice). Hilford et al. (2002) reported a series of experiments and found that although source recognition ROCs were clearly curvilinear, they were nevertheless not as curvilinear as the UVSD model predicts. Note that the UVSD model that applies to associative and source memory tests is, in its essentials, the same as the one that applies to item recognition. As illustrated in Figure 3, the models are characterized by two Gaussian distributions and a decision criterion. Only

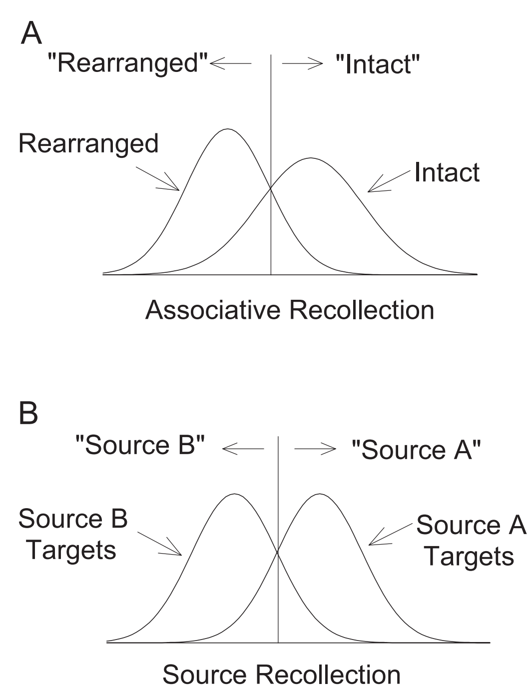

Figure 3. Idealized signal-detection models for associative recognition (Figure 3A) and source recognition (Figure 3B).

the nature of the decision axis changes. In the source memory case, for example, the decision variable represents the degree of source recollection in favor of one source or the other (i.e., recollection is construed as a continuous variable, not as a threshold variable). Because the models are otherwise just like the model that applies to old–new recognition, a curvilinear ROC should be observed in each case, and they should be accurately described by the UVSD model whether memory strength is weak or strong.

What are the theoretical implications of the systematic deviations from the ROC predictions derived from signal-detection theory for associative and source memory? Kelley and Wixted (2001) and Hilford et al. (2002) independently arrived at very similar explanations in their studies of associative recognition and source memory. Specifically, both assumed that on some proportion of trials, associative information (or source information) was not available. Kelley and Wixted allowed for the possibility that the occasional absence of associative information implied a retrieval threshold, but Hilford et al. offered an even simpler explanation, namely, that source information is occasionally unavailable simply because participants do not always successfully encode this information during study. It is easy to imagine how that might happen given that in both procedures, item information must be successfully processed before the associative or source information can be encoded along with it.

If associative or source information is not encoded on some trials, then the relevant detection model would be a mixture model. That is, in addition to the typical Source A and Source B distributions, there would be a third strength distribution (midway between the other two) that corresponds to the cases in which no source information was encoded. Hilford et al. (2002) showed that the presence of that third distribution would introduce just the anomalies in the ROC curve that they observed, and DeCarlo (2003) provided further evidence of the validity of this idea.

If that explanation is correct, then if one simply allows for more study time during encoding (which would decrease the likelihood that no associative or source information is encoded), the ROC distortion should disappear. As indicated earlier, Kelley and Wixted (2001) found that strengthening word pairs on the associative recognition task did indeed eliminate the anomalies that had been observed in the associative recognition ROC. Figure 4 shows the associative recognition ROC from the weak and strong conditions from three experiments of that study. The curvilinear detection model and the linear all-or-none recollection model rival each other in the weak condition. That is, the ROC in that condition, although curvilinear, is more linear than it should be according to the UVSD account. However, in the strong condition—which is a condition in which recollection presumably plays a much stronger role—the UVSD model fits almost perfectly, whereas the linear DPSD model exhibits large and systematic deviations. If recollection does predominate in the strong condition, the DPSD model would anticipate a more linear ROC in that condition (not the curvilinear ROC predicted by the UVSD account).

Slotnick and Dodson (2005) recently offered what would appear to be the decisive test of the ability of the UVSD and DPSD models to accommodate source memory ROCs. Using a standard source memory procedure in which items were presented in a male voice or a female voice, they showed that as is typically true, the source ROC was curvilinear but was also more linear than it should have been according to the UVSD model. To test whether

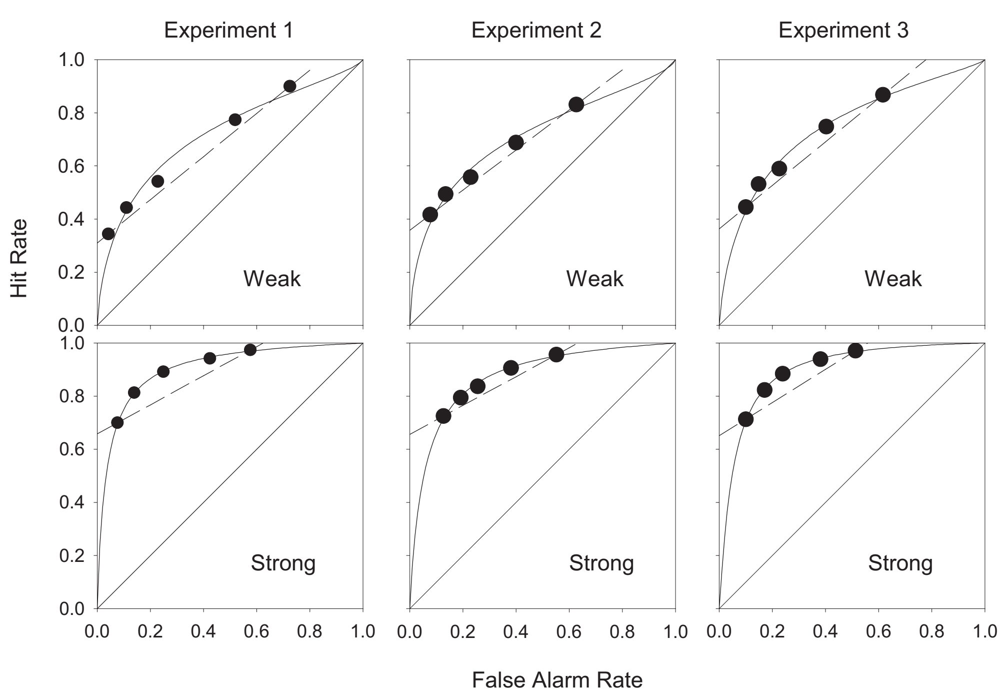

Figure 4. Associative recognition receiver operating characteristic (ROC) data for the weak and strong conditions of three experiments reported by Kelley and Wixted (2001). The solid curve in each ROC shows the best fitting unequal-variance signal-detection model, and the dashed line shows the best fitting threshold model.

this linear distortion resulted from the inclusion of items for which no source information was encoded, Slotnick and Dodson constructed source ROCs for only those items that initially received a relatively high-confidence old decision. The UVSD and DPSD models make clear predictions about what should happen when the source ROC is constructed this way. The DPSD model assumes that high-confidence old decisions are largely based on recollection. In fact, this model assumes that recollection leads exclusively to a high-confidence old decision, so an especially high percentage of these responses will be based on recollection (e.g., see Figure 7b in Yonelinas, 2001). As such, the DPSD model clearly predicts that constructing a source ROC from high-confidence old responses should yield a linear ROC. The dual-process UVSD model proposed by Wixted and Stretch (2004) and Rotello et al. (2004) also assumes that items that receive a high-confidence old decision are the most likely to be associated with source information and, critically, are also the least likely to be items for which no source information is available. However, because the UVSD model assumes that recollection is a continuously distributed variable, it predicts that this approach should eliminate the usual ROC distortion and produce a classic curvilinear ROC. By contrast, the DPSD model predicts that this approach should simply raise the intercept of the linear ROC. The differing predictions are stark, so the test is an especially decisive one. Slotnick and Dodson reported findings that strongly supported the UVSD account. That is, source ROCs constructed from items that received high-confidence (and largely recollection-based) old decisions were free of the linear distortion and were accurately characterized by the UVSD model. This was even true of the linear source memory ROC initially reported by Yonelinas (1999a). When those data were reanalyzed by constructing a source ROC from items that initially received a high-confidence old decision (i.e., when it was constructed mainly from items associated with a high degree of recollection), the ROC was clearly curvilinear. Slotnick and Dodson concluded that "source ROCs can appear linear when nondiagnostic source information is included in the analysis" (p. 151). They further concluded that "the unequal variance model accounted for both recognition memory and source memory ROCs, supporting a continuous process of memory retrieval" (p. 151). These results confirm a suggestion first made by Ratcliff, McKoon, and Tindall (1994), which is that the inclusion of guessing responses can distort the shape of the ROC (cf. Malmberg & Xu, 2006).

Taken together, these findings show that the evidence that once legitimized the DPSD model—and seemed to pose problems for the UVSD model—told an incomplete story. A considerable amount of recent research has shown that ROCs from associative recognition and source memory studies are typically curvilinear—as the UVSD model requires—and are reliably better fit by the UVSD model than by the DPSD model. Further, this research has determined why, in weak memory conditions, the ROC often appears to be more linear than it should be. The linearity is not a reflection of a threshold recollection process, as originally supposed, but instead reflects the fact that under weak memory conditions, associative and source information is sometimes not encoded. The crucial test, performed by Slotnick and Dodson (2005), was to construct a source memory ROC that both models would agree was composed largely of recollection-based decisions. That ROC was well characterized by the UVSD model, not the DPSD model.

Non-ROC Evidence That Recollection Is a Continuous Process

The curvilinear ROCs observed in associative recognition and source memory procedures suggest that recollection is not a discontinuous threshold process that either supports a high-confidence old decision or plays no role at all, but is instead a continuously distributed process that is associated with varying degrees of confidence. Indeed, independent (non-ROC) evidence on that issue points to the same conclusion. In their classic article on source monitoring, Johnson et al. (1993) noted the following:

> However, in Jacoby's work, recollection of prior episodes has often been treated as something that either occurs or does not occur. Research on source monitoring, in contrast, has emphasized the idea that recollection occurs in degrees and produces variations in the phenomenal qualities of memories. (p. 13)

The evidence supporting this claim is abundant, and evidence against it is hard to come by. Dodson, Holland, and Shimamura (1998), for example, presented words on a study list in two different male voices and two different female voices and found that participants often remembered partial detail about the source (e.g., that the word was spoken in a female voice) even when the specific source (Female 1 or Female 2) could not be remembered. Simons, Dodson, Bell, and Schacter (2004) recently reported the same phenomenon and further showed that when young and older adults were matched in overall recognition, age-related deficits were observed for both specific- and partial-source recollection. Evidence of all-or-none recollection was not observed in either study. Although these findings are suggestive, one might imagine that even partial recollection of the voice always engendered high confidence that the item was previously encountered, in which case the core assumption of the DPSD model—that a high-confidence decision is made whenever recollection of any kind occurs—would still be viable.

The well-known tip-of-the-tongue (TOT) phenomenon also implies that recollection comes in degrees. The TOT state is one in which the sought-after information is not accessible even though one is sure that it exists in memory and that its accessibility is imminent (Brown & McNeill, 1966; Schwartz, 2002). As Maril, Simons, Weaver, and Schacter (2005) put it, "As such, the TOT represents an intermediate level of recall success in which the sought-after information is not recalled, but at the same time, it is distinctly different from a retrieval failure" (p. 1130). They also summarized extensive evidence suggesting that the TOT state involves partial recollection in which only fragments of the original encoding experience are retrieved. Studies have shown, for example, that although an item cannot be recalled, limited information about that item often can be, such as the first and last letters of the sought-after word (Koriat & Lieblich, 1974), the number of syllables in the word (Rubin, 1975), the syntactic properties of sentences (Miozzo & Caramazza, 1997; Vigliocco, Vinson, Martin, & Garrett, 1999), and similar words that are related in sound or meaning (Cohen & Faulkner, 1986).

Findings like these speak to the continuous nature of recollection, but the TOT state, by definition, involves confidence that the item will be ultimately recalled. However, it is easy to imagine that even lesser degrees of recollection (and lower degrees of confidence) can occur. In fact, as indicated next, recollection can be sufficiently impoverished that participants fail to appreciate that they are recollecting partial information even though that information supports above-chance responding. That scenario seems quite incompatible with the notion that partial retrieval always engenders a high state of confidence.

Koriat, Levy-Sadot, Edry, and de Marcas (2003) investigated this issue by asking what people know about words that cannot be completely recalled. Participants who were fluent in Hebrew first learned the Hebrew translations of pseudo-Somali words. Following a retention interval, a cued-recall test was given with the Somali word serving as the cue. Even when participants denied knowing the corresponding Hebrew word and denied knowing anything about its semantic attributes, they were nevertheless asked to judge whether the unrecalled word was good or bad, strong or weak, active or passive—the three dimensions of the semantic differential (Osgood, 1952). An independent group of participants had previously rated the Hebrew words along these dimensions, and their ratings served as the basis to determine what the experimental participants knew about the words they could not recall. The results showed that they knew much about those words because their ratings along the various dimensions of the semantic differential for the nonrecalled words were significantly above chance levels. This was true even though participants were not confident enough to indicate that they knew anything about those words. Koriat et al. summarized their results in the following way:

> The results of the present study clearly indicate that information about the semantic attributes of a word may be accessible even when the word itself cannot be recalled. These results are consistent with the view that memory and forgetting are not all-or-nothing processes. (p. 1102)

Koriat et al. (2003) also asked for remember–know judgments both for words that were recalled and for words that could not be recalled but were assigned ratings on the semantic differential. Their summary of those results is interesting as well:

> Note that not only was item recall sometimes associated with know responses but also attribute information was sometimes associated with remember responses. This pattern is consistent with recent evidence (e.g., Hicks, Marsh, & Ritschel, 2002; see also Conway & Dewhurst, 1995) suggesting that source monitoring is not an all-or-nothing process, and that even vague, partial information can support source monitoring. (Koriat et al., 2003, p. 1102)

The idea that vague, partial information can support source monitoring seems inconsistent with the way in which recollection operates in the DPSD model, according to which any degree of recollection yields a high-confidence decision. It is, however, completely consistent with the source ROC study reported by Slotnick and Dodson (2005), which showed that source memory ROCs, which are based on the full range of confidence ratings, are curvilinear and are accurately characterized by the UVSD model.

In summary, a large body of evidence suggests that recollection is a continuous variable that can involve degrees of strength (and varying degrees of confidence) ranging from low to high. Curvilinear associative recognition and source memory ROCs clearly point to this conclusion, and the non-ROC research discussed earlier points to the same conclusion. These findings would appear to largely undermine the assumptions of the DPSD model, according to which recollection always occasions a high-confidence decision, whereas decisions made with lower confidence are based purely on familiarity. It is that (apparently incorrect) assumption about the nature of recollection that allows the model to be used to extract estimates of recollection and familiarity from ROC data. However, if the model's assumptions about recollection are incorrect, then its estimates of recollection and familiarity are not likely to be valid.

All of these considerations lead to an important and testable prediction concerning the effect of partial source recollection on confidence in the initial old–new recognition decision. According to the DPSD model, source information that is recollected during the old–new decision stage (e.g., recollecting that the word was presented in a female voice) leads to a high-confidence old decision. But if recollection is a graded phenomenon, then a different pattern should be observed. Specifically, source information recollected during the old–new decision stage should lead to old decisions that are made with varying degrees of confidence, depending on how much source information is retrieved. This issue, which is more critical that it might appear to be at first glance, is discussed next in relation to a DPSD model.

An Alternative DPSD Theory

A considerable body of evidence supports the idea that recognition memory is governed by two processes, recollection and familiarity (e.g., Mandler, 1980), and the preceding review shows that a considerable body of evidence supports the UVSD account of recognition memory as well. In the past, the two models have existed independently and even seemed to be at odds with each other. However, once one accepts the possibility that recollection is a continuous variable that is associated with varying degrees of confidence, then these two long-standing models are easily reconciled. As Wixted and Stretch (2004) and Rotello et al. (2004) proposed, a model that is compatible with both theories holds that old–new recognition decisions are made with increasing confidence to the extent that graded recollection of source and associative information—combined with familiarity—is retrieved. In other words, the unidimensional memory strength variable of signal-detection theory consists of the additive combination of two processes, recollection and familiarity.

In the simplest version of this model, the baseline strengths of the items are assumed be normally distributed as $N(\mu_b, \sigma^2_b)$, where μ and σ^2 denote mean and variance, respectively, and the subscript *b* stands for baseline. Targets can be conceived of as random variables drawn from that distribution that have had strength added from two processes, recollection (*r*) and familiarity (*f*). If these added strengths are distributed as $N(\mu_r, \sigma^2_r)$ and $N(\mu_f, \sigma^2_f)$, respectively, then the mean of the target distribution will be $\mu_b + \mu_r + \mu_f$, and the variance will be $\sigma^2_b + \sigma^2_r + \sigma^2_f$ (assuming that baseline strength, recollective strength, and familiarity are all independent). Thus, the target distribution will have a greater mean and greater variance than the lure distribution. If recollection and familiarity are correlated across items, which seems likely, and if the added strength is independent of baseline strength, then the mean and variance of the target distribution would be $\mu_b + \mu_r + \mu_f$ and $\sigma^2_b + (\sigma^2_r + \sigma^2_f + 2\rho\sigma_r\sigma_f)$, respectively, where ρ is the correlation coefficient. This simple model merely assumes that the memory strength of each target item is jointly determined by recollection and familiarity, which are both added to the item's baseline strength at study.[1]

According to this view, the memory strength variable is not ahistorical. Instead, it is comprised of multiple sources of information, including recollective information. In their classic critique of strength theory, Anderson and Bower (1972) noted that participants are capable of many fine distinctions (e.g., list discrimination) that are not captured by a simple, ahistorical view of memory strength. As they put it, "Example differentia in verbal learning experiments would be *where* in space the item was presented, *who* said it, and *how* it was said" (p. 99). In the dual-process UVSD model proposed by Wixted and Stretch (2004), all of this information is assumed to be included into the continuously distributed memory strength signal in a manner analogous to that of a juror incorporating multiple sources of evidence into an overall assessment of a defendant's guilt. To conceptualize the juror's assessment as a strength variable is not to deny its heterogeneous nature, and the same is true of the memory strength variable.

In the past, efforts to reconcile dual-process theory with signal-detection theory have left intact a basic assumption common to all dual-process models (namely, that individual recognition decisions are based either on recollection or on familiarity), followed by some kind of modification in signal-detection theory (e.g., Atkinson & Juola, 1973; Yonelinas, 1994). By dropping that either/or assumption, dual-process theory is naturally accommodated by signal-detection theory. In fact, as proposed by Wixted and Stretch (2004), dropping the either/or assumption seems sensible in light of the graded nature of the recollection process. If recollection were an all-or-none variable, such that its occurrence would always support a high-confidence old decision, then it would make sense to base a recognition decision solely on that process whenever recollection occurred (without any regard for familiarity) and to resort to familiarity as a backup whenever recollection failed to occur. However, the logic changes the moment one accepts the possibility that recollection is a graded phenomenon that is associated with lower degrees of confidence as well. Why would a decision be based exclusively on recollection if only some information about the word is recollected? And why would a decision be based solely on familiarity if partial recollection for that item happens to be available as well? If both recollection and familiarity are continuous variables, combining them into a single memory signal would make more sense than responding based on either one alone. And if they are combined into a single memory strength variable, then the UVSD model naturally applies (except that now it is a dual-process UVSD model). Moreover, under this scenario, a clear prediction emerges about the relationship between confidence in an old–new recognition decision and the accuracy of a subsequent recollection-based decision, as discussed next.

Relationship Between Old–New Recognition and Source Accuracy

In a source memory procedure, participants are typically asked for an old–new decision and then asked to indicate the item's source. According to the DPSD model, source information that is recollected during the old–new decision stage leads to a high-confidence old response. If recollection fails, then the old–new decision is based on familiarity and is usually made with lower confidence. Thus, high-confidence old decisions should be followed by accurate source decisions, whereas less confident old decisions (which are theoretically based on familiarity) should be followed by source decisions that are no better than chance. If source information is instead graded, as the findings of Dodson et al. (1998) and Koriat et al. (2003) and many others suggest, then it should be the case that confidence judgments on an initial old–new recognition test will be related in a graded fashion (not in an all-or-none fashion) to the accuracy of subsequent source decisions. That is, for the initial old–new decision, any source information that is retrieved should increase confidence that the item was seen before on the list. If a moderate amount of source information is retrieved, then the old judgment might be made with moderate confidence (and the subsequent source discrimination would be made with moderate accuracy, on average).

Yonelinas (2001) reported a striking pattern of results that appeared to offer strong support for the assumptions that underlie the DPSD model. The results were taken from a source memory experiment in which participants first made an old–new decision and then made a source judgment (indicating whether the item had been presented in a male voice or a female voice) for each item. Old–new confidence ratings were made on a 6-point scale (1 = *high confidence that the item was new*; 6 = *high confidence that the item was old*). As predicted by the DPSD model, source accuracy was above chance only for items that received an initial old–new confidence rating of 6 (see Yonelinas, 2001, Figure 7b, which summarizes the results of his Experiment 3). This is as it should be according to the DPSD model.

However, the same issue was addressed in great detail by Slotnick et al. (2000) and Slotnick and Dodson (2005), although their data were presented in detailed fashion, making the trends of interest here somewhat hard to discern. Table 3 shows their results in a more condensed fashion (i.e., collapsed over source confidence ratings) to reveal the critical information. The data show an unmistakable trend: As confidence in the old–new recognition decision increased, accuracy in the subsequent source decision increased as well. Note that Table 3 shows confidence in the old–new decision (not confidence in the source decision) and accuracy in the subsequent source decision (not accuracy in the old–new decision). What these data show is that in graded fashion, increasing confidence in the old–new decision is associated with increasing accuracy in the subsequent source decision. And although this pattern was not observed in Experiment 3 of Yonelinas (1999a), it was observed in his Experiment 2. A reanalysis of those data shows that source accuracy scores for items that received high, medium, and low confidence ratings were 76%, 58%, and 52% correct, respectively. Findings like these weigh against the DPSD model as a viable model of recognition performance, and they support the dual-process UVSD model proposed by Wixted and Stretch (2004) and Rotello et al. (2004).

[1] Gaussian distributions are used here for the sake of simplicity, but it seems more sensible to assume that the distributions are non-Gaussian (e.g., the lower bound of the Gaussian is $-\infty$ even though the lower bound of memory strength should be 0). In practice, the continuous Weibull distribution, which is bounded at zero and skewed to the right, fits ROC data as well as the Gaussian distribution. However, the conclusions drawn from fitting that distribution (and using it to illustrate recollection and familiarity) are the same as those drawn by using the more familiar Gaussian distribution.

Table 3
Source Accuracy as a Function of Confidence in the Old–New Recognition Decision

Old–new confidence	Slotnick et al. (2000) Exp. 2	Slotnick et al. (2000) Exp. 3	Slotnick and Dodson (2005) Exp. 1	Slotnick and Dodson (2005) Exp. 2	*M*
High	.840	.862	.802	.816	.830
Medium	.672	.620	.617	.618	.632
Low	.653	.562	.528	.518	.565

Note. Exp = experiment. Data are from "An Analysis of Signal Detection and Threshold Models of Source Memory," by S. D. Slotnick, S. A. Klein, C. S. Dodson, and A. P. Shinamura, 2000, *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* pp. 1508 and 1512, and from "Support for a Continuous (Single-Process) Model of Recognition Memory and Source Memory," by S. D. Slotnick and C. S. Dodson, 2005, *Memory & Cognition, 33,* p. 163.

Remembering and Knowing

Exactly the same issues arise with regard to the use of remember–know judgments. As they are ordinarily interpreted, remember–know judgments fit naturally with the DPSD model. That is, theoretically, a remember response is supplied when recollection occurs (in an essentially all-or-none fashion). A know response is supplied when recollection fails and the decision is based exclusively on familiarity. Estimates of recollection and familiarity derived from the remember–know procedure are often said to agree with estimates derived from fitting the DPSD model to confidence-based ROC data (Yonelinas, 2002), although Rotello, Macmillan, Reeder, and Wong (2005) recently reported findings suggesting that the recollection estimates derived from the two procedures do not converge.

The considerations discussed above with regard to confidence in old–new recognition decisions and subsequent accuracy in source judgments can be brought to bear on the issue of what remember responses and know responses measure. In a typical remember–know experiment, participants are presented with a list of words and are later asked to decide whether the words are old or new. For each word that is declared to be old, they are further asked to indicate if they remember its appearance on the list or just know that it appeared on the list. If this procedure is used in a source-monitoring experiment, one can ask about the accuracy of source recollection associated with remember and know judgments. According to the standard dual-process interpretation, know responses represent familiarity-based memories that are devoid of recollection. As such, source decisions that follow know judgments should be at chance levels. Conway and Dewhurst (1995) and Perfect, Mayes, Downes, and Van Eijk (1996) investigated this issue. The former study found that know judgments were indeed associated with greater-than-chance source accuracy, whereas the latter study found that to be true in only one out of three experiments. Still, as noted by Hintzman (2001), the two negative experiments did yield above-chance source accuracy for know decisions even though the effect was not significant. It therefore seems that know judgments, like medium-confidence old–new decisions, are associated with some degree of source recollection.

In a recent neuroimaging study by Eldridge, Engel, Zeineh, Bookheimer, & Knowlton (2005), a remember–know procedure was used in a source-monitoring experiment. Picture items were presented in one of four colors and in one of two orientations that could be asked about after participants indicated that the item was old and made a remember–know decision. Although they interpreted their imaging results as if remember responses reflected recollection and know responses reflected familiarity, they also reported the probability that a participant would recollect an encoding detail (color or orientation) as a function of whether the old–new decision was remember or know. These values, which were estimated from their Figure 1, are shown in Table 4. Eldridge et al. found that recollection success was significantly above chance levels for know responses (but below that for remember responses). In other words, know responses involved less recollective detail than remember responses but did not signal the absence of recollective detail.

Once again, these findings are inconsistent with the idea that know responses reflect familiarity-based decisions, and they weigh against the closely related DPSD model. How do they bear on the UVSD model? The UVSD interpretation of remember–know judgments has been widely debated in recent years. Donaldson (1996) argued that remember and know responses might reflect different degrees of memory strength instead of qualitatively different memory processes, which is another way of saying that these judgments can be understood within the standard UVSD framework. According to this account, participants adopt two decision criteria, one of which is situated at a relatively low point on the decision axis (the know criterion) and the other of which is situated at a relatively high point on the decision axis (the remember criterion). This idea is illustrated in Figure 5. If the detection interpretation is correct, then remember–know judgments do not reflect qualitatively different forms of memory, and results based on these judgments cannot be used to investigate dual-process theories of recognition memory. Instead, when memory strength is high, participants supply a remember response (cf. Benjamin, 2005; Dunn, 2004; Rotello et al., 2005). When memory strength is lower, they supply a know response. And according to the dual-process UVSD model advanced by Wixted and Stretch (2004), low memory strength is associated with lower degrees of recollection and familiarity (not a strong familiarity process that is devoid of recollection). As such, source accuracy for know decisions should be greater than chance (but below that associated with remember responses).

As Perfect et al. (1996) noted, if know responses are associated with above-chance recollective accuracy "then serious questions

Table 4
Source Accuracy as a Function of Remember and Know Judgments for the Old–New Recognition Decision

Judgment	Color	Orientation
Chance	.25	.50
Know	.31*	.57*
Remember	.47*	.69*

Note. Data are from "A Dissociation of Encoding and Retrieval Processes in the Human Hippocampus," by L. L. Eldridge, S. A. Engel, M. M. Zeineh, S. Y. Bookheimer, and B. J. Knowlton, 2005, *Journal of Neuroscience, 25,* p. 3282. Copyright 2005 by the Society for Neuroscience. Adapted with permission.
* $p < .05$ (significantly greater than chance).

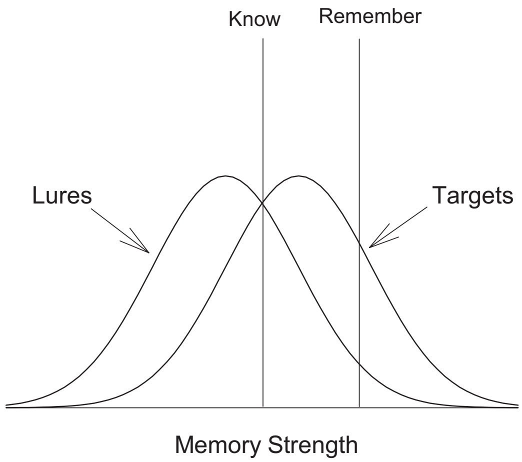

Figure 5. Signal-detection interpretation of remember–know judgments.

would be raised about the suitability of the remember/know procedure as a means of exploring the hypothetical underlying explicit and implicit memory processes" (p. 799). Although know responses being associated with above-chance recollective accuracy would raise serious questions about the remember–know procedure as it is commonly understood, it is just what Wixted and Stretch's (2004) model predicts, and the available data would appear to support that prediction.

These considerations fit naturally with a study conducted by Hicks, Marsh, and Ritschel (2002). Participants first made a source decision and then made a remember–know judgment about that decision. Source accuracy was very high in this experiment (close to 90% correct), which, from a recollection/threshold perspective, means that many items exceeded the recollection threshold. As such, if remember responses capture above-threshold recollection, one would expect most of these source decisions to be given remember judgments. Instead, the responses were almost equally likely to receive remember and know judgments. Stated another way, a very high percentage of know judgments were accompanied by accurate source information, which should not happen if know responses reflect familiarity-based responding.

The overall point to be made is that the literature is almost unanimous in its rejection of the notion that recollection is an all-or-none process that leads to a high-confidence old decision or to a remember judgment. Instead, the literature strongly suggests that recollection is a graded phenomenon that is associated with low-confidence old–new decisions and with know judgments as well. That pattern is compatible with the UVSD model but not with the DPSD model (at least not as that model has been used to date).

In Search of the Neural Correlates of Recollection and Familiarity

The considerations discussed above have important implications for investigations into the neural correlates of the component processes of recognition memory. The most important implication is that recognition memory itself is not process pure (i.e., individual recognition decisions are not based on one process or the other). Instead, it seems that the memory strength axis in the standard UVSD model represents aggregate memory strength (i.e., recollection and familiarity together). If so, asking for a remember–know judgment and requiring that the participant choose one or the other is not disentangling recollection and familiarity because the two processes are entwined in both cases. Both recollection and familiarity are high in strength, on average, when a remember response is given and are low in strength, on average, when a know response is given. For the same reason, fitting the DPSD model to ROC data cannot disentangle those processes even though doing this yields parameter estimates as if it can. Yet, as described next, conclusions about the neural substrates of recognition memory are often based on these models.

ROC Analysis in Lesion Studies

A relatively new strategy in the neuroscience literature is to compare ROC data produced by controls with ROC data produced by individuals (or rats) with hippocampal lesions. More specifically, the ROCs are fit by the DPSD model to extract estimates of familiarity and recollection, and the estimates are then compared across groups to determine if the recollection process is selectively impaired in the lesion group. Yonelinas et al. (2002) found that, as predicted by a model that assumes that the hippocampus selectively subserves recollection (Aggleton & Shaw, 1996), control participants exhibited the typical asymmetrical curvilinear ROC, one with a *z*-ROC slope of less than 1.0 (which the DPSD model takes as the signature of recollection). By contrast, participants with amnesia—who were assumed to have hippocampal lesions—produced a symmetrical ROC with a *z*-ROC slope of approximately 1.0, which in this model is the signature of pure familiarity-based responding. The results were taken as evidence that the hippocampus exclusively supports the recollection process and plays no role in the familiarity process (which is theoretically governed by adjacent brain structures along the parahippocampal gyrus).

Because the UVSD model has been strongly supported over the DPSD model whenever they have been directly compared, it stands to reason that these ROC results should be interpreted in terms of the former model, not the latter. What do these findings imply when interpreted in terms of the UVSD model? The results suggest that control participants—who have stronger memories—yield data that are consistent with the UVSD model, whereas individuals with amnesia—who have weaker memories—yield data that are consistent with the equal-variance signal-detection model. Such a result is not surprising in light of evidence showing that as memory strength weakens, the difference in variance between the target and lure distributions often decreases as well (Glanzer et al., 1999). This result is not invariably observed (e.g., Ratcliff et al., 1992), but as Shiffrin and Steyvers (1997) noted, "For subjects receiving only moderate training or less, the conditions that produce higher accuracy usually produce a lower NRS [slope of the *z*-ROC]" (p. 150). Because participants with amnesia produce lower accuracy than control participants, it seems reasonable to suppose that those with amnesia might exhibit a more symmetrical ROC for that reason alone.

Recently, Fortin et al. (2004) conducted an odor recognition study with rats in which ROCs were generated with a reinforcement-biasing manipulation. Control rats produced a typical asymmetrical curvilinear ROC when recognition was tested following a 30-min interval. By contrast, rats with hippocampal lesions tested under the same conditions exhibited weaker memory and produced a symmetrical curvilinear ROC. This pattern matches the pattern observed in humans (Yonelinas et al., 1998, 2002) and is what might be expected according to the UVSD model because weaker memory conditions often yield more symmetrical ROCs (Glanzer et al., 1999). However, Fortin et al. also weakened the memories of control rats by introducing a long (70 min) retention interval, which resulted in a low level of recognition performance that was similar to that of the hippocampal rats. Even so, the ROC produced following a long retention interval was not symmetrical and curvilinear but was instead nearly linear. According to the DPSD model, a linear ROC occurs when responding is exclusively based on recollection. Fortin et al. interpreted their data in terms of this model, so they argued that responding was purely recollection based (presumably because familiarity faded rapidly to zero as the retention interval increased).

These findings are noteworthy for several reasons. First, the linear ROC appears to be the first one ever reported for an old–new recognition procedure in the history of experimental psychology. Whereas linear ROCs have been reported for associative recognition and source memory procedures, no other linear ROCs can be identified in the much larger literature involving the old–new procedure. In fact, the absence of linear ROCs for old–new recognition explains why the field so thoroughly rejected the original high-threshold theory in years gone by. Further, prior claims regarding linear ROCs for associative recognition and source memory have not withstood scrutiny. Accordingly, a cautious attitude toward this newly reported linear ROC seems warranted. Still, the results of this study, had they been presented as a test of the UVSD model versus the DPSD model, would have to be taken as differentially supporting the latter. As always, the UVSD account must struggle to explain a linear ROC, whereas it is easily reconciled with a model that assumes all-or-none recollection.

Then again, given the large number of studies that have supported the UVSD model over the DPSD model in humans, it makes sense to ask whether some other variable (other than all-or-none odor recollection) may have accounted for these surprising results. The reason a linear ROC was found is not clear, but one possibility is that the biasing manipulation that was used played a role. To induce the changes in bias needed to produce an ROC in rats, reinforcer outcomes for correct responses to targets and to lures were varied in a variety of ways (e.g., the outcomes differed in food amounts and in the difficulty of acquiring the food). An ideal ROC procedure would manipulate bias without affecting memory strength, but the use of differential reward outcomes is known to sometimes affect memory strength as well as bias in animals. If memory strength was affected by their biasing manipulations in the long-retention-interval condition (such that memory was stronger when the reinforcing outcomes were more asymmetrical), then the shape of the ROC would be affected.

Whatever the explanation of their linear ROC, a recent study by Wais, Wixted, Hopkins, and Squire (2006) suggests that the same result is not observed in humans. In this study, young adults were randomly assigned to one of five retention interval conditions (1 hr, 1 day, 1 week, 2 weeks, and 8 weeks). The recognition performance of the young adults decayed as expected over time, with performance following the 8-week retention interval being only slightly (but significantly) greater than chance. Of importance, the ROCs at each retention interval were clearly curvilinear. Even at the longest retention interval, where the ROC must become more linear as it approaches the diagonal, the data were much better described by a symmetrical curvilinear model than by a linear recollection model. In addition, the ROC was typically asymmetrical at the 1-hr retention interval and became ever more symmetrical as the overall level of performance decreased. These retention interval results are consistent with what Glanzer et al. (1999) reported for a variety of other strength manipulations (e.g., study time, list length, word frequency), but they are inconsistent with the results observed in rats (Fortin et al., 2004). It is not clear why rats yielded such an unusual ROC, but the point is that the same result has not been observed in humans.

The fact that the young adults in Wais et al.'s (2006) study yielded an increasingly symmetrical ROC as memory strength weakened is consistent with the idea that individuals with amnesia tend to exhibit symmetrical ROCs because they have weak memories (not because their hippocampal lesions have selectively destroyed the recollection process). If this way of thinking is correct, then if memory in those with amnesia were somehow strengthened (e.g., by extra study time or by using shorter study lists), then they too should exhibit an ROC asymmetry. To test this prediction, Wais et al. tested 6 participants with amnesia with damage restricted to the hippocampus and 8 age-matched older controls, all of whom studied 50-item lists and were tested following a 3-min retention interval. The participants with hippocampal lesions were then tested again wtih shorter (10-item) lists, which brought the level of their recognition memory performance closer to that of the elderly controls. This was the critical test.

The ROC produced by the older control participants following a 3-min retention interval was typically asymmetrical. The ROC produced by the participants with amnesia tested under the same conditions was symmetrical. However, when the participants with amnesia were tested with shorter lists, thereby increasing their memory performance to a level near that of the older controls, their ROC was as asymmetrical as that of the older controls. This should not have happened (a) if the asymmetry is a reflection of recollection, as the DPSD model assumes, and (b) if the hippocampus subserves the recollection process. Instead, the results suggest that the processes that determine the shape of the ROC are operative in those with hippocampal lesions and in control participants. What differs is the overall strength of memory in the two groups, where strength is a joint function of recollection and familiarity.

The results reported by Wais et al. (2006) are consistent with the idea that both the hippocampus and adjacent parahippocampal structures subserve both recollection and familiarity. Additional compelling support for this idea was recently reported by Rutishauser, Mamelak, and Schuman (2006), who recorded firing rates from individual hippocampal neurons, using electrodes that had been implanted into the brains of people who had had surgery for epilepsy. A source memory procedure was used in which the participants with epilepsy were presented with pictures in one of four quadrants on the screen. On the recognition test, the pictures were presented at the center of the screen for an old–new recognition decision, followed by a source memory test for the picture's original location.

Rutishauser et al. (2006) found that some neurons in the hippocampus showed altered firing rates only during the second presentation of the picture stimulus (i.e., only when the item appeared on the recognition test), reflecting memory for the stimulus. This pattern was observed even for trials in which the source recollection decision was incorrect and even for participants whose source recollection performance was at chance. This pattern suggests that these hippocampal neurons were firing on the basis of item familiarity, which is contrary to the idea that the hippocampus selectively subserves recollection. If the hippocampus supports both recollection and familiarity, and if hippocampal activity is a primary determinant of memory strength, then it seems reasonable to suppose that the memory strength of a particular test item is determined by both processes (as the UVSD model proposed here assumes). This is not a necessary assumption, but the idea that recollection and familiarity are both subserved by the hippocampus seems well-suited to the idea that both processes contribute to the memory strength of an individual test item.

Neuroimaging

Similar considerations apply to the neuroimaging literature. Yonelinas, Otten, Shaw, and Rugg (2005) relied on the DPSD model to investigate the neural correlates of recollection and familiarity. Participants first studied a list of 150 words and were then presented with a recognition test in which the test items were presented one at a time for a recognition decision. For each word, they were asked to select one of five options to indicate how they remembered that word as having been presented in the study phase. The five response options were 1, 2, 3, 4, and R. If they were able to recollect something specific about seeing the word at study, participants were asked to choose R (for remember). If they could not recollect anything specific about experiencing the item, they were asked to rate their memory confidence from low to high (1, 2, 3, or 4). Thus, this was a variant of the widely used remember–know procedure except that, instead of allowing participants to say "know" for items that were declared to be old in the absence of recollection, they were asked to express varying degrees of confidence in their ostensibly familiarity-based decisions. Note that this procedure—like the typical remember–know procedure—assumes that individual recognition decisions are based either on one process or on the other.

The neural correlates of recognition memory retrieval were examined by functional magnetic resonance imaging (fMRI), and activity associated with R responses (relative to activity associated with confidence ratings of 4) was taken to reflect recollection, whereas activity that correlated with confidence ratings from 1 to 4 was taken to reflect degrees of familiarity. Different areas of activity were associated with these two kinds of ratings across a variety of brain regions. For example, a lateral parietal/temporal region was related to R responses, and a more superior parietal region was related to confidence ratings from 1 to 4. The former was assumed to be involved in recollection and the latter with familiarity. Similarly, in medial parietal regions, the posterior cingulate was theoretically related to recollection, whereas the precuneus was theoretically related to familiarity. Results like these were taken to mean that recollection and familiarity are subserved by different networks in the brain.

This interpretation assumes that the DPSD model is correct and—by implication—that the UVSD model is wrong. Again, however, a large body of evidence suggests that it is the other way around. In particular, it seems very likely that varying degrees of confidence are associated with varying degrees of recollection (not just varying degrees of familiarity) and that participants recollect information about an item even when they do not realize it. What, then, accounts for the observed dissociation? According to the UVSD account, remember responses reflect high-strength, high-confidence memories (not memories uniquely associated with recollection). However, it seems reasonable to suppose that remember responses are also more likely than know responses to be associated with the subjective awareness that episodic information is being recollected. Tulving (1983) proposed a special state of subjective consciousness called *autonoetic awareness,* according to which one experiences a sense of mental time travel back to the original encoding event. In fact, the remember response category was designed to identify that experience.

Unlike recollection, which occurs in degrees, autonoetic awareness may be a more threshold-like process. If so, then the neural signature of autonoetic awareness (not the neural signature of recollection) may be what Yonelinas et al. (2005) identified as correlating with activity in the posterior cingulate and other regions of the brain. Because memories that exceed the remember criterion are strong memories, and because only strong memories are likely to engender autonoetic awareness, neural activity associated with autonoetic awareness shows up as activity correlated with remember responses. However, the key point is that such activity may track a subjective state, not recollection itself.

An explanation along these lines offers one possible interpretation of remember–know dissociations that is consistent with the UVSD model in that it does not assume that the type of memory associated with remember responses is qualitatively different from the type of memory associated with know responses (even though the type of subjective experience may often be).

Whereas a distinct neural signature associated with remember responses may reflect the conscious awareness of recollection (not the exclusive presence of recollection), other imaging results may simply reflect the fact that remember responses denote stronger memories than know responses. Thus, especially in the medial temporal lobe, one might expect to find graded memory strength effects associated with remember–know judgments (because activity in the medial temporal lobe is often thought to be a determinant of memory strength). For example, Eldridge et al. (2000) found that activity in the hippocampus was higher, relative to baseline, only for hits that were accompanied by a remember response. Hits that were accompanied by a know response yielded activity comparable with that observed during quiet baseline periods. Because some have argued that the hippocampus is involved mainly in recollection (not familiarity), results such as these appear to validate the assumption that remember responses are process-pure indicators of recollective success. However, a simpler possibility—one that is entirely consistent with the UVSD account—is that the increased brain activity associated with remember responses reflects the higher level of activity that is associated with the retrieval of strong memories. The retrieval of weaker memories that give rise to a know response may not have yielded detectable hippocampal activity because that activity was measured with respect to a baseline period, one that was free of experimenter-imposed activity but that likely involved a considerable amount of encoding and retrieval anyway. Stark and Squire (2001), for example, found that activity in the medial temporal lobe was higher during quiet rest periods than during tedious tasks, such as deciding whether numbers were odd or even, and concluded that "rest is apparently an active condition associated with significant cognitive activity" (p. 12765). Activity during "rest" may, in fact, be as high as that associated with the retrieval of the relatively weak memories that give rise to know responses.

A recent neuroimaging study by Gonsalves, Kahn, Curran, Norman, and Wagner (2005) helps to drive home some of these points. Using both fMRI and the magnetoencephalography, they found that activity in the anterior medial temporal lobe at retrieval was inversely correlated with overall memory strength. Although they used the remember–know procedure, they also collected confidence ratings, and they found what Wixted and Stretch (2004) had previously shown to be true, namely, that despite instructions encouraging participants to respond remember or know only when confidence was high (leading many researchers to mistakenly believe that know responses reflect high-confidence, familiarity-based responses), confidence tended to be high in remember responses and low in know responses. As Gonsalves et al. put it,

> Within this framework, recent evidence suggests that, in contrast to common assumptions, R/K decisions are not necessarily process pure. Rather, R decisions can be based on high levels of recollection and also can be associated with high levels of familiarity, such that R/K responses bear a systematic relation with gradations in recognition confidence (Wixted & Stretch, 2004). . . . Consistent with the interpretation that R decisions may reflect high levels of familiarity (most likely together with recollection), the present one-step behavioral expressions of remembering and knowing mapped to different points in the recognition confidence continuum, with R responses being predominantly associated with the upper two recognition confidence levels and K responses being predominantly associated with less confident old responses (Figure 1B). (p. 757)

Moreover, their conclusion that activity in the medial temporal cortex reflects a continuously varying memory signal that is perceived as varying degrees of memory strength is entirely compatible with the UVSD model. Findings like these were previously reported in a meta-analysis by Henson, Cansino, Herron, Robb, and Rugg (2003), but they took the decreasing activity as a function of memory strength to be a familiarity signal. The results reported by Gonsalves et al. (2005) suggest that it might be better thought of as a memory strength signal, a signal that is probably composed of recollection as well as familiarity.

General Discussion

The search for the neural correlates of recognition memory is not, and never can be, a theory-free endeavor. Nearly a half century of recognition memory research has yielded one theoretical account of recognition memory decision making—signal-detection theory—that has guided thinking more effectively than any other, including its predecessor (high-threshold theory) and, according to the arguments presented here, its presumed successor (namely, the DPSD model). The DPSD model is an elegant, parsimonious, and internally consistent attempt to bring together the dual-process theory of recognition memory and signal-detection theory. However, its weakness is that it regards recollection as a threshold phenomenon. In so doing, it changes the essential character of the UVSD model. Although that assumption allows the DPSD model to extract estimates of recollection and familiarity from the ROC, this same assumption renders it unable to compete with UVSD theory in describing ROC data. A model that provides a consistently inferior description of ROC data—whether those data are derived from an old–new recognition procedure, an associative recognition procedure, or a source memory procedure—cannot reasonably be used to estimate the component processes of recognition memory from the ROC. The fact that it provides a consistently poorer fit suggests that its underlying assumptions are incorrect.

In a recent case study involving a patient with hippocampal damage, Cipolotti et al. (2006) reported that the DPSD model provided a good fit to the ROC data produced by the patient and the control participants. They further remarked that this "suggests that our estimates of recollection and familiarity give reliable values of the contribution of these two processes" (Cipolotti et al., 2006, p. 502). The allure of a good fit is powerful, but the main point made by Roberts and Pashler (2000) was that a good fit, in and of itself, does not provide compelling evidence in favor of a model. In spite of its good fit, the fact that a different model (the UVSD model) consistently offers a better fit suggests that the DPSD model is based on incorrect assumptions and, therefore, that the estimates of recollection and familiarity it provides are not valid.

A dual-process model that is true to the assumptions of signal-detection theory simply holds that the component processes (recollection and familiarity) are both continuous variables that are additively combined into a strength-of-memory signal. Kelley and Wixted (2001) first proposed such a model in the context of an associative recognition procedure, and Wixted and Stretch (2004) and Rotello et al. (2004) proposed that the same idea applies to old–new recognition.[2] The most important implication of this way of thinking is that individual recognition memory decisions are not process pure, so efforts to measure the pure processes of recollection and familiarity with the remember–know procedure or ROC analysis are misplaced.

It seems reasonable to suppose that recognition decisions are not process pure in light of compelling evidence suggesting that the recollection process, like the familiarity process, is a graded phenomenon. If both processes are continuous, such that both can be associated with varying degrees of confidence, then not combining them into a single memory signal would be like a juror who does not combine multiple sources of evidence when assessing the defendant's degree of guilt. Although an assessment of guilt could be based on one piece of evidence or the other (e.g., either fingerprint evidence or fiber evidence), combining multiple indicators into the overall assessment would be more efficient, and the same holds true for an assessment of memory strength.

This is not to say that it is impossible to maintain the process-purity assumption even if recollection is assumed to be a graded process. For example, one could assume that a recognition decision is based exclusively on recollection whenever the graded recollection signal exceeds a criterion strength (i.e., whenever recollection is strong enough) and that otherwise the decision is based exclusively on familiarity. Murdock (in press) and Diana, Reder, Arndt, and Park (2006) have recently advanced models

[2] Rotello et al. (2004) offer a different account of remember–know judgments, but their account of old–new decision making is the same as that suggested by Wixted and Stretch (2004) so long as one interprets specific strength and global strength to refer to recollection and familiarity.

along these lines. The decision-making aspects of these process models are similar to the dual-process UVSD model proposed by Wixted and Stretch (2004) except that they assume that the two criteria (essentially the remember and know criteria) operate on two signal distributions—a recollection distribution and a familiarity distribution—instead of one. Both models allow for continuously distributed memory strength distributions even for recollection, but their ability to adequately characterize ROC data has not yet been established. However, Rotello et al. (in press) found that a similar variant of the DPSD model did not describe ROC data as well as the UVSD model. In addition, these models do not offer any compelling reason why decisions would be based on either one process or the other, even though they are both graded variables that are imperfect indicators of prior occurrence (and so would be more usefully combined into a single signal).

One Process or Two?

It has long been assumed that signal-detection theory is inherently incompatible with a dual-process account of recognition memory because the unidimensional memory strength axis upon which the target and lure distributions are placed seems to imply one process. However, as indicated earlier, continuously distributed recollection and familiarity processes can be additively combined into a single, unidimensional memory strength variable, which means the UVSD model is compatible with dual-process theory after all. Although it is compatible with dual-process theory, it seems important to recognize that the UVSD model is compatible with a single-process interpretation as well. If one were to advance a single-process view, however, it might be best to avoid labeling the decision axis as *familiarity*, because the evidence suggests that varying degrees of confidence in the old–new decision are associated with varying degrees of recollection (cf. Anderson & Bower, 1972).

If ROC analyses are compatible with a single-process interpretation, why assume that two processes underlie recognition memory judgments? The main reason is that a considerable body of non-ROC evidence supports the idea that two processes are involved. Mandler (1980), for example, summarized research showing that the degree to which a list is semantically organized (which theoretically influences the recollection process only) affects slow recognition responses but not fast ones (although Gillund & Shiffrin, 1984, did not find evidence like this). Mandler also described compelling anecdotal evidence for a familiarity process, such as when one is absolutely sure of having encountered a person in the past even though no information about that person can be retrieved. Also, Jacoby (1991) showed that the variables that affect one's ability to tell whether an item has been recently encountered (old vs. new) differ from the variables that affect one's ability to identify which of two lists the item appeared on (which presumably involves recollection). Hintzman, Caulton, and Levitin (1998) further showed that the retrieval dynamics associated with old–new decisions differ qualitatively from the retrieval dynamics associated with list discrimination.

At a minimum, it seems clear that information about the recent occurrence of an item (familiarity) differs from the retrieval of information about its contextual detail (recollection). Whether familiarity reflects the early products of a retrieval process that unfolds over time (and eventually includes recollective detail) or whether it reflects a separate process—one that obeys different rules—remains a matter of debate. The main point to emphasize here is that the UVSD model is (a) strongly supported by recent work on the form of the ROC and (b) applies whether the memory strength variable is thought to arise from a single process or from the additive combination of two or more continuous processes.

A Signal-Detection-Based Approach to Neuroimaging

The considerations discussed above have important implications for how neuroimaging studies are conducted. Because individual old–new recognition decisions are often assumed to be process pure, a preliminary goal of many imaging investigations is to separate the subset of old–new decisions that are based on recollection from those that are based on familiarity so that medial temporal lobe activity for each kind of decision can be separately contrasted with a baseline condition. Imaging studies often rely on the remember–know procedure for this purpose (e.g., Eldridge et al., 2005; Eldridge, Knowlton, Furmanski, Bookheimer, & Engel, 2000). However, if individual old–new decisions are not process pure, then another approach to understanding the neural basis of recognition memory is needed.

What are the implications for neuroimaging if the UVSD model provides an accurate account of recognition memory? One implication is that it makes sense to search for a correlation between *degree of memory strength* along a decision axis and *degree of neural activity* in a particular region of the brain. The use of the remember–know procedure may accomplish this goal to some extent because, according to the signal-detection interpretation, remember responses denote high-strength memories and know responses denote low-strength memories. Thus, although they are ordinarily taken to reflect recollection-based and familiarity-based decisions, neuroimaging studies that rely on this technique are more likely to shed light on the neural correlates of strong and weak memories, with strong memories involving an abundance of recollection and relatively high levels of familiarity as well (cf. Gonsalves et al., 2005).

Although remember–know judgments could be used to measure the neural correlates of memory strength, confidence ratings would provide a much better approach. First, participants are intimately familiar with confidence ratings, and there is some evidence that participants' lack of familiarity with remember–know judgments (which, unlike confidence ratings, require detailed instructions) leads to substantial item-to-item variation in criterion placement (Rotello et al., in press; Wixted & Stretch, 2004). If so, that only adds noise to an already noisy situation. Second, the remember–know procedure, even if it is taken to reflect degrees of memory strength, does not take advantage of gradations of strength associated with items that are judged to be new. The confidence ratings associated with new decisions are as lawful as those associated with old decisions (and are in accordance with detection theory). For example, as with confidence in old decisions, accuracy for confidence in new decisions is high when confidence is high and low when confidence is low (Ratcliff & Murdock, 1976).

How can confidence ratings be used to investigate questions of interest in the neuroimaging literature? Consider the question of the neural correlates of item versus source recollection. To identify the neural correlates of source memory, a common strategy involves drawing a contrast between (a) correctly identified old items that also have correct source decisions and (b) correctly identified old items that are followed by incorrect source decisions (e.g., Davachi, Mitchell, & Wagner, 2003). A possible problem with this approach is that the presence of source information is confounded with confidence in the initial old decision. That is, old decisions that are followed by correct source judgments are typically made with much higher confidence, on average, than items that are subsequently followed by incorrect source decisions (Slotnick & Dodson, 2005; Slotnick et al., 2000). Moreover, the strategy does not take advantage of the graded nature of recollection.

An alternative strategy to investigating the neural correlates of source recollection is suggested by results reported by Slotnick et al. (2000) and Slotnick and Dodson (2005). They showed that ROC anomalies in source recognition derive from the inclusion of items for which no source information is available. Those no-information items can be excluded by constructing the source ROC only for items that initially received a relatively high-confidence old decision (which also serves to eliminate the confound discussed above). The source ROCs constructed with this method are curvilinear (consistent with source recollection being a matter of degree) and are well characterized by the UVSD model. As such, if one's goal is to identify the neural correlates of source recollection, it makes sense to correlate neural activity with the degree of confidence in source recollection for items that initially received a high-confidence old decision. For such items, confidence and accuracy in source judgments vary continuously over a wide range (i.e., high-confidence source decisions are associated with high accuracy, and low-confidence source decisions are associated with low accuracy). Also, it seems reasonable to suppose that the items are all highly familiar and that what varies is the degree of source information that is recollected. If the hippocampus subserves the graded recollection process, one might hypothesize that hippocampal activity at retrieval will correlate with the degree of confidence (and, therefore, with the degree of accuracy) in the source decision.

Similar considerations apply when the goal is to identify the neural correlates of item information independent of source information. A common strategy is to contrast correctly identified targets without source information (i.e., targets that are correctly declared to be old but are then incorrectly attributed to the wrong source) versus incorrectly identified target items (i.e., targets that are incorrectly declared to be new). A possible problem with this approach is that it tends to exclude high-confidence item decisions (because those tend to be associated with correct source decisions).

How might the full range of item strength be used without introducing a confound with the presence of source information? Low-confidence source decisions are usually associated with source accuracy that is not much above chance (e.g., Slotnick & Dodson, 2005; Slotnick et al., 2000). However, the old–new item decisions that precede those low-confidence source decisions are made with the full range of confidence (low to high) despite the relative absence of source information. Conceivably, variations in confidence for these items reflect variations in item familiarity, although one cannot rule out the possibility that recollection other than source recollection (e.g., recollection of the item's position in a list) is playing a role. Still, a reasonable question to ask is whether activity in particular regions of the brain, such as the hippocampus, correlate with varying degrees of confidence (and varying degrees of accuracy) in the item decision when source information is unavailable. This approach is similar to that used by Rutishauser et al. (2006), who assessed the neural basis of familiarity by recording the activity of individual hippocampal neurons for old items when source memory performance was at chance.

The overarching strategy, then, is to base imaging analyses on methods that yield clean behavioral ROC data that are well characterized by a longstanding psychological model that has been differentially supported in the recent literature (namely, the UVSD model). Imaging studies with confidence ratings have been conducted occasionally, but they have been conducted without reference to or in explicit opposition to a signal-detection perspective (e.g., Chua, Rand-Giovannetti, Schacter, Albert, & Sperling, 2004; Chua, Schacter, Rand-Giovannetti, & Sperling, 2006). Obviously, a neuroimaging strategy based on signal-detection considerations is not guaranteed to yield useful information about the neural correlates of recognition memory. However, as a general rule, it seems more sensible to investigate the neural basis of memory with guidance from a well-defined and empirically supported psychological theory than to use methods that are based mainly on intuition or on models that have been largely disconfirmed in the psychological literature.

A Signal-Detection-Based Alternative to the Search for Process Purity

The dual-process theory of recognition memory has spawned decades of research based on the idea that individual recognition decisions are process pure. However, the signal-detection approach suggests a related but nevertheless distinctly different avenue of investigation. In a detection-based approach, the emphasis would be on the *dimensions* of recognition memory instead of (only) on its component processes.

Recognition memory judgments can be made along more than one dimension, not just the old–new dimension. Indeed, the old–new dimension may be a derivative dimension that is based on several more basic dimensions of memory. The basic or primary dimensions of memory are those that are psychologically independent of (i.e., orthogonal to) each other. Orthogonal dimensions are such that the strengths along those dimensions for a given item are not correlated with each other (such that high strength along one dimension does not imply high strength along the other). Although strengths along those primary dimensions are not correlated with each other, they may all be correlated with strength along the derivative old–new dimension.

The primary dimensions of recognition memory. What are the primary dimensions of recognition memory? The answer to that question is not known, but some candidates can be identified. Banks (1970, 2000) was the first to clearly stipulate that the source memory axis was different from the old–new axis, and his claim has been reinforced by Hilford et al. (2002), by Glanzer et al. (2004), and more recently, by Slotnick and Dodson (2005). All of these researchers illustrated the difference between old–new and source memory dimensions, using a figure like the one shown in Figure 6 (which is Figure 1 from Slotnick & Dodson, 2005). Conceivably, the source memory axis, unlike the old–new axis, can be thought of as a primary dimension that is independent of other primary dimensions. In addition to the source memory axis, recent evidence suggests that one's sense of *recency* (or judgment of recency [JOR]) and one's sense of *frequency* (or judgment of frequency [JOF]) may be primary dimensions that are independent

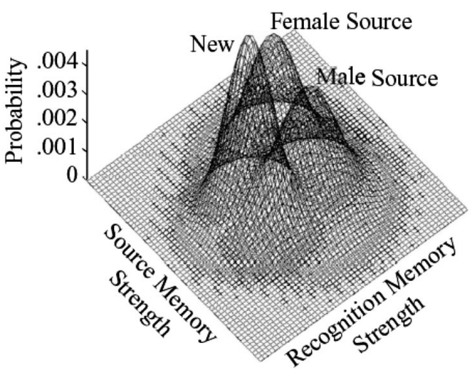

Figure 6. Two dimensional signal-detection model for an old–new recognition and source recognition task (in which the sources are a male voice and a female voice). From "Support for a Continuous (Single-Process) Model of Recognition Memory and Source Memory," by S. D. Slotnick and C. S. Dodson, 2005, *Memory & Cognition, 33,* p. 152. Copyright 2005 by the Psychonomic Society. Reprinted with permission.

of each other (and possibly independent of the source memory axis as well). Hintzman (2001) argued that his findings were consistent with the model shown in Figure 7 (which is Figure 8 from Hintzman, 2001). The diagonal axis is the old–new decision axis (and the remember and know criteria along that axis are also shown). The JOR axis and the JOF axis are different from the old–new axis and, more to the point, are different from and possibly *orthogonal* to each other. As Hintzman (2001) observed, one simple reason to believe that these dimensions are orthogonal is that "subjects ordinarily show little tendency to confuse recency and frequency" (p. 1354). Hintzman further noted that

> the same unidimensional strength cannot be the primary basis of both JOF and JOR. Frequency and recency are different aspects of experience and are remembered as such. JOR is affected only minimally by repetition (Flexser & Bower, 1974; Peterson, 1967). Likewise, JOF is affected little, if at all, by within-list differences in recency (Galbraith & Underwood, 1973; Hintzman, 1969), and subjects can parcel out the overall experimental frequencies of words to different temporally defined lists, with considerable accuracy (Hintzman & Block, 1971). (p. 1353)

If JOFs are minimally affected by how recently the item was seen, and if JORs are minimally affected by how frequently an item was seen, then it cannot be the case that JOF and JOR both represent pure measures of familiarity.

Although the JOR and JOF axes may be largely orthogonal to each other, neither dimension is likely to be orthogonal to the old–new decision axis in Hintzman's (2001) model. That is, as either sense of recency or sense of frequency increases, strength of evidence along the old–new axis increases as well (i.e., the participant becomes more confident that the item is old). In an interesting observation that is relevant to the present discussion, Hintzman, echoing Underwood (1969), suggested that strength along the old–new axis "is not a primary property of memory, but a 'byproduct of the attributes.'. . . Although the temporal and frequency attributes are primary, they are not the only dimensions relevant to memory-based decisions" (p. 1355). That is, as indicated earlier, memory strength along the old–new decision axis is derivative from more basic dimensions (i.e., more basic attributes), such as frequency, recency, and source memory, all of which may be orthogonal to each other.

Establishing mnemonic independence. Although very little work has been conducted on the issue of orthogonal decision axes in recognition memory, an exactly analogous issue has been addressed in the literature on object perception. The question of *perceptual independence* asks whether one aspect of perception (e.g., size) is independent of another (e.g., weight). The technique that has been developed to address that question is derived from general recognition theory (Ashby & Townsend, 1986) and involves holding the physical stimulus constant (e.g., a small cylinder) and testing whether trial-to-trial variations in the perception of one dimension (e.g., size estimates) are correlated with trial-to-trial variations in the perception of the other (e.g., weight estimates). When confidence ratings are taken along both dimensions, as might be done in a memory experiment, Ashby (1988) showed that the correlation between the two sets of ratings can be estimated with the tetrachoric correlation instead of the Pearson *r*. Wickens and Olzak (1989, 1992) described a more direct way to fit a multidimensional UVSD model to the data to estimate the degree of independence. This approach is very similar to the standard method of fitting a UVSD model to ROC data. To fit a standard ROC, two significant parameters are adjusted to maximize the likelihood of the data, with one parameter representing discriminability (e.g., d_e) and the other representing the ratio of the standard deviations of the target and lure distributions. If data from two dimensions are being fit, each dimension would be associated with two parameters, and an additional parameter would capture the degree of correlation between the two. That parameter is the parameter of interest (and it should be close to zero if the dimensions are independent). Even more powerful distribution-free tests of independence are described by Ashby and Townsend (1986).

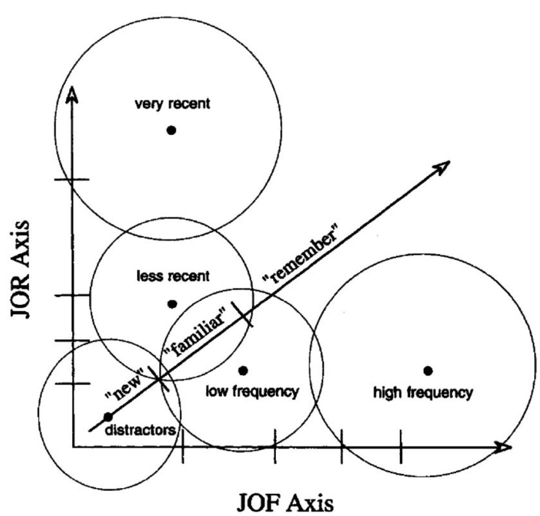

Figure 7. Two dimensional signal-detection model for a judgment of recency (JOR) task and a judgment of frequency (JOF) task. From "Judgments of Frequency and Recency: How They Relate to Reports of Subjective Awareness," by D. L. Hintzman, 2001, *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* p. 1355. Copyright 2001 by the American Psychological Association. Reprinted with permission.

The question of perceptual independence is very much like what I refer to as the question of *mnemonic independence.* Mnemonic independence can be assessed by conducting an experiment involving any two recognition memory decision axes, but the general technique is described for an experiment involving decisions made along the *source memory axis* and the *sense of recency* (or JOR) axis. The correlation between recency judgments from a particular recency condition (e.g., lag of 16) and confidence ratings in source (separately for items that were presented in a male voice and a female voice) can be assessed with the methods described above. If the correlation is typically positive across participants, it would suggest that one's sense of recency is not independent of one's confidence that the item was heard in a particular voice. But if these are independent dimensions, as prior research suggests, then the correlation should be zero.

Independence versus separability. Perceptual independence (and mnemonic independence) should be distinguished from the related phenomenon of *perceptual separability* (and, by analogy, *mnemonic separability*). Perceptual separability is said to exist when the *physical value* of one perceptual component does not affect the perception of the other. For example, if the actual size of an object did not affect the perceived weight of the object, then perceptual separability would hold.[3] However, researchers know this is violated because participants typically experience the size–weight illusion (i.e., the actual size of an object affects the perception of weight even with weight held constant). Mnemonic separability seems equally unlikely to hold. That is, it seems almost certain that a recently presented item will be judged to have been presented recently *and* will be associated with accurate source memory (because the memory trace will be relatively strong), whereas an item presented longer ago will be judged to have been presented less recently and will be associated with less accurate source memory (because the memory trace will be relatively weak). Note that separability is concerned with the relationship between the physical value of the stimulus along one dimension (e.g., its actual recency) and the perception of that stimulus along another (e.g., confidence in the source judgment). The more interesting question concerns mnemonic independence, which concerns the relationship between perceived memory strength along one dimension and perceived memory strength along another. For a particular recency condition (e.g., the item was presented 10 min ago), the participant's perception of recency will exhibit trial-to-trial variability. The question is whether items that for whatever reason, generate a relatively strong sense of recency also generate a relatively high-confidence source decision. If not, then the recency and source dimensions are independent.

Establishing the primary, psychologically independent dimensions of recognition memory is a task that has not yet been undertaken by the field of experimental psychology. However, it seems like a natural avenue to pursue in light of compelling evidence suggesting that signal-detection theory offers the most viable account of recognition memory confidence judgments. If independent dimensions are eventually established, then searching for the neural correlates of confidence along those psychologically independent dimensions would seem to be a reasonable next step. That is, it seems reasonable to suppose that psychologically independent dimensions of memory are subserved by different neural structures. Identifying the brain structures that underlie different dimension of memory might be accomplished by correlating neural activity (measured by fMRI) with strength along each dimension.

Conclusion

The main point of this article is that signal-detection theory is much more viable than alternative theories that currently guide investigations into the brain basis of recognition memory, such as the DPSD model and the theory that remember–know judgments capture recollection and familiarity. However, this is not meant to imply that signal-detection theory provides a complete account of recognition memory. Some ROC findings are not easily accommodated by signal-detection theory (e.g., Van Zandt, 2000), but they are no more easily handled by the DPSD model. Also, the theory does not specify the computational mechanisms that give rise to recollection and familiarity, and it cannot handle the full range of reaction time data obtained in recognition memory experiments (Ratcliff & Murdock, 1976). These limitations are shared by the DPSD model, and it seems clear that a more complete account, such as that offered by the diffusion model (e.g., Ratcliff, Thapar, & McKoon, 2004), will be needed to explain the full pattern of reaction time data. That said, it is worth noting that the diffusion model envisions a continuous process of information accumulation, which makes it more compatible with the continuous UVSD model than with the dichotomous DPSD model.

The search for the neuroanatomical basis of memory must be guided by some psychological theory—ideally, one that is conceptually simple, comprehensive, and empirically flawless. Until such a theory becomes available, however, it seems reasonable to suggest that the guidance should be provided by the most viable psychological theory that is also simple enough to offer heuristic value, and that theory would appear to be the UVSD model.

[3] If the physical change in one stimulus does not affect accuracy on the other over a range of values, then perceptual separability probably holds. However, there is some chance that, even under these conditions, perceptual separability is violated. If a physical change in Stimulus A equally affects the perception of Stimulus B regardless of the physical value of Stimulus B, then a violation of perceptual separability would not be detected.

References

Aggleton, J. P., & Shaw, C. (1996). Amnesia and recognition memory: A reanalysis of psychometric data. *Neuropsychologia, 34,* 51–62.

Aggleton, J. P., Vann, S. D., Denby, C., Dix, S., Mayes, A. R., Roberts, N., & Yonelinas, A. P. (2005). Sparing of the familiarity component of recognition memory in a patient with hippocampal pathology. *Neuropsychologia, 43,* 1810–1823.

Anderson, J. R., & Bower, G. H. (1972). Recognition and retrieval processes in free recall. *Psychological Review, 79,* 97–123.

Ashby, F. G. (1988). Estimating the parameters of multidimensional signal detection theory from simultaneous ratings on separate stimulus components. *Perception & Psychophysics, 44,* 195–204.

Ashby, F. G., & Townsend, J. T. (1986). Varieties of perceptual independence. *Psychological Review, 93,* 154–179.

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: 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: Freeman.

Banks, W. P. (1970). Signal detection theory and human memory. *Psychological Bulletin, 14,* 81–99.

Banks, W. P. (2000). Recognition and source memory as multivariate decision processes. *Psychological Science, 11,* 267–273.

Benjamin, A. S. (2005). Recognition memory and introspective remember/know judgments: Evidence for the influence of distractor plausibility on "remembering" and a caution about purportedly nonparametric measures. *Memory & Cognition, 33,* 261–269.

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

Brown, M. W., & Aggleton, J. P. (2001). Recognition memory: What are the roles of the perirhinal cortex and hippocampus? *Nature Reviews Neuroscience, 2,* 51–61.

Brown, R., & McNeill, D. (1966). The "tip of the tongue" phenomenon. *Journal of Verbal Learning and Verbal Behavior, 5,* 325–337.

Cary, M., & Reder, L. M. (2003). A dual-process account of the list-length and strength-based mirror effects in recognition. *Journal of Memory and Language, 49,* 231–248.

Chua, E. F., Rand-Giovannetti, E., Schacter, D. L., Albert, M. S., & Sperling, R. A. (2004). Dissociating confidence and accuracy: fMRI shows origins of the subjective memory experience. *Journal of Cognitive Neuroscience, 16,* 1131–1142.

Chua, E. F., Schacter, D. L., Rand-Giovannetti, E., & Sperling, R. A. (2006). Understanding metamemory: Neural correlates of the cognitive process and subjective level of confidence in recognition memory. *NeuroImage, 29,* 1150–1160.

Cipolotti, L., Bird, C., Good, T., Macmanus, D., Rudge, P., & Shallice, T. (2006). Recollection and familiarity in dense hippocampal amnesia: A case study. *Neuropsychologia, 44,* 489–506.

Cohen, G., & Faulkner, D. (1986). Memory for proper names: Age differences in retrieval. *British Journal of Developmental Psychology, 4,* 187–197.

Conway, M. A., & Dewhurst, S. A. (1995). Remembering, familiarity, and source monitoring. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 48*(A), 125–140.

Davachi, L., Mitchell, J. P., & Wagner, A. D. (2003). Multiple routes to memory: Distinct medial temporal lobe processes build item and source memories. *Proceedings of the National Academy of Sciences, 100,* 2157–2162.

DeCarlo, L. T. (2002). Signal detection theory with finite mixture distributions: Theoretical developments with applications to recognition memory. *Psychological Review, 109,* 710–721.

DeCarlo, L. T. (2003). An application of signal detection theory with finite mixture distributions to source discrimination. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29,* 767–778.

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.

Dodson, C. S., Holland, P. W., & Shimamura, A. P. (1998). On the recollection of specific- and partial-source information. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24,* 1121–1136.

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

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

Egan, J. P. (1958). *Recognition memory and the operating characteristic* (Tech. Note AFCRC-TN-58–51). Bloomington: Indiana University, Hearing and Communication Laboratory.

Egan, J. P. (1975). *Signal detection theory and ROC analysis.* New York: Academic Press.

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.

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

Flexser, A. J., & Bower, G. H. (1974). How frequency affects recency judgments: A model for recency discrimination. *Journal of Experimental Psychology, 103,* 706–716.

Fortin, N. J., Wright, S. P., & Eichenbaum, H. (2004). Recollection-like memory retrieval in rats is dependent on the hippocampus. *Nature, 431,* 188–191.

Galbraith, R. C., & Underwood, B. J. (1973). Perceived frequency of concrete and abstract words. *Memory & Cognition, 1,* 56–60.

Gardiner, J. M., Ramponi, C., & Richardson-Klavehn, A. (2002). Recognition memory and decision processes: A meta-analysis of remember, know and guess responses. *Memory, 10,* 83–98.

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

Glanzer, M., Hilford, A., & Kim, K. (2004). Six regularities of source recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 1176–1195.

Glanzer, M., Kim, K., Hilford, A., & Adams, J. K. (1999). Slope of the receiver-operating characteristic in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 500–513.

Gonsalves, B. D., Kahn, I., Curran, T., Norman, K. A., & Wagner, A. D. (2005). Memory strength and repetition suppression: Multimodal imaging of medial temporal cortical contributions to recognition. *Neuron, 47,* 751–761.

Green, D. M., & Swets J. A. (1966). *Signal detection theory and psychophysics.* New York: Wiley.

Healy, M. R., Light, L. L., & Chung, C. (2005). Dual-process models of associative recognition in young and older adults: Evidence from receiver operating characteristics. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31,* 768–788.

Heathcote, A. (2003). Item recognition memory and the ROC. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29,* 1210–1230.

Henson, R. N. A., Cansino, S., Herron, J. E., Robb, W. G. K., & Rugg, M. D. (2003). A familiarity signal in human anterior medial temporal cortex. *Hippocampus, 13,* 259–262.

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.

Hilford, A., Glanzer, M., Kim, K., & DeCarlo, L. T. (2002). Regularities of source recognition: ROC analysis. *Journal of Experimental Psychology: General, 131,* 494–510.

Hintzman, D. L. (1969). Apparent frequency as a function of frequency and the spacing of repetitions. *Journal of Experimental Psychology, 80,* 139–145.

Hintzman, D. L. (2001). Judgments of frequency and recency: How they relate to reports of subjective awareness. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* 1347–1358.

Hintzman, D. L., & Block, R. A. (1971). Repetition and memory: Evidence for a multiple-trace hypothesis. *Journal of Experimental Psychology, 88,* 297–306.

Hintzman, D. L., Caulton, D. A., & Levitin, D. J. (1998). Retrieval dynamics in recognition and list discrimination: Further evidence of separate processes of familiarity and recall. *Memory & Cognition, 26,* 449–462.

Hintzman, D. L., & Curran, T. (1994). Retrieval dynamics of recognition and frequency judgments: Evidence for separate processes of familiarity and recall. *Journal of Memory and Language, 33,* 1–18.

Hockley, W. E. (1991). Recognition memory for item and associative information: A comparison of forgetting rates. In W. E. Hockley & S. Lewandowsky (Eds.), *Relating theory and data: Essays on human memory in honor of Bennet B. Murdock* (pp. 227–248). Hillsdale, NJ: Erlbaum.

Howard, M. W., Bessette-Symons, B. A., Zhang, Y., & Hoyer, W. J. (2006). Aging selectively impairs recollection in recognition memory for pictures: Evidence from modeling and ROC curves. *Psychology and Aging, 21,* 96–106.

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

Jacoby, L. L., & Dallas, M. (1981). On the relationship between autobiographical memory and perceptual learning. *Journal of Experimental Psychology: General, 3,* 306–340.

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin, 114,* 3–28.

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

Koriat, A., Levy-Sadot, R., Edry, E., & de Marcas, G. (2003). What do we know about what we cannot remember? Accessing the semantic attributes of words that cannot be recalled. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29,* 1095–1105.

Koriat, A., & Lieblich, I. (1974). What does a person in the "TOT" state know that a person in a "don't know" state does not know? *Memory & Cognition, 2,* 647–655.

Kroll, N. E. A., Yonelinas, A. P., Dobbins, I. G., & Frederick, C. M. (2002). Separating sensitivity from response bias: Implications of comparisons of yes–no and forced-choice tests for models and measures of recognition memory. *Journal of Experimental Psychology: General, 131,* 241–254.

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide* (2nd ed.). Mahwah, NJ: Erlbaum.

Malmberg, K. J., & Xu, J. (2006). The influence of averaging and noisy decision strategies on the recognition memory ROC. *Psychonomic Bulletin & Review, 13,* 99–105.

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

Maril, A., Simons, J. S., Weaver, J. J., & Schacter, D. L. (2005). Graded recall success: An event-related fMRI comparison of tip of the tongue and feeling of knowing. *Neuroimage, 24,* 1130–1138.

Miozzo, M., & Caramazza, A. (1997). Retrieval of lexical–syntactic features in tip-of-the-tongue states. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 1410–1423.

Murdock, B. (in press). Decision-making models of remember/know judgments. *Psychological Review.*

Osgood, C. E. (1952). The nature and measurement of meaning. *Psychological Bulletin, 49,* 197–237.

Perfect, T. J., Mayes, A. R., Downes, J. J., & Van Eijk, R. (1996). Does context discriminate recollection from familiarity in recognition memory? *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 49*(A), 797–813.

Peterson, L. R. (1967). Search and judgment in memory. In B. Kleinmuntz (Ed.), *Concepts and the structure of memory* (pp. 153–180). New York: Wiley.

Pitt, M. A., Kim, W., & Myung, I. J. (2003). Flexibility versus generalizability in model selection. *Psychonomic Bulletin & Review, 10,* 29–44.

Qin, J., Raye, C. L., Johnson, M. K., & Mitchell, K. J. (2001). Source ROCs are (typically) curvilinear: Comment on Yonelinas (1999). *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* 1110–1115.

Ratcliff, R., McKoon, G., & Tindall, M. (1994). Empirical generality of data from recognition memory receiver-operating characteristic functions and implications for the global memory models. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 763–785.

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

Ratcliff, R., Sheu, C. F., & Gronlund, S. D. (1992). Testing global memory models using ROC curves. *Psychological Review, 99,* 518–535.

Ratcliff, R., Thapar, A., & McKoon, G. (2004). A diffusion model analysis of the effects of aging on recognition memory. *Journal of Memory and Language, 50,* 408–424.

Roberts, S., & Pashler, H. (2000). How persuasive is a good fit? A comment on theory testing. *Psychological Review, 107,* 358–367.

Rotello, C. M., Macmillan, N. A., Hicks, J. L., & Hautus, M. (in press). Interpreting the effects of response bias on remember–know judgments using signal-detection and threshold models. *Memory & Cognition.*

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., Macmillan, N. A., & Van Tassel, G. (2000). Recall-to-reject in recognition: Evidence from ROC curves. *Journal of Memory and Language, 43,* 67–88.

Rubin, D. C. (1975). Within word structure in the tip-of-the-tongue phenomenon. *Journal of Verbal Learning & Verbal Behavior, 14,* 392–397.

Rutishauser, U., Mamelak, A. N., & Schuman, E. N. (2006). Single-trial learning of novel stimuli by individual neurons of the human hippocampus–amygdala complex. *Neuron, 49,* 805–813.

Schwartz, B. L. (2002). *Tip-of-the-tongue states: Phenomenology, mechanism, and lexical retrieval.* Hillsdale, NJ: Erlbaum.

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

Simons, J. S., Dodson, C. S., Bell, D., & Schacter, D. L. (2004). Specific and partial source memory: Effects of aging. *Psychology and Aging, 19,* 689–694.

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.

Slotnick, S. D., Klein, S. A., Dodson, C. S., & Shimamura, A. P. (2000). An analysis of signal detection and threshold models of source memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 1499–1517.

Smith, D. G., & Duncan, M. J. J. (2004). Testing theories of recognition memory by predicting performance across paradigms. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 615–625.

Stark, C. E. L., & Squire, L. R. (2001). When zero is not zero: The problem of ambiguous baseline conditions in fMRI. *Proceedings of the National Academy of Sciences of the USA, 98,* 12760–12766.

Tulving, E. (1983). *Elements of episodic memory.* London: Oxford University Press.

Underwood, B. J. (1969). Attributes of memory. *Psychological Review, 76,* 559–573.

Van Zandt, T. (2000). ROC curves and confidence judgments in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 582–600.

Verde, M. F., & Rotello, C. M. (2004). Strong memories obscure weak memories in associative recognition. *Psychonomic Bulletin & Review, 11,* 1062–1066.

Vigliocco, G., Vinson, D. P., Martin, R. C., & Garrett, M. F. (1999). Is "count" and "mass" information available when the noun is not? An investigation of tip of the tongue states and anomia. *Journal of Memory and Language, 40,* 534–558.

Wais, P. E., Wixted, J. T., Hopkins, R. O., & Squire, L. R. (2006). The hippocampus supports both the recollection and the familiarity components of recognition memory. *Neuron, 49,* 459–466.

Wickens, T. D., & Olzak, L. A. (1989). The statistical analysis of concurrent detection ratings. *Perception & Psychophysics, 45,* 514–528.

Wickens, T. D., & Olzak, L. A. (1992). Three views of association in concurrent detection ratings. In F. G. Ashby (Ed.), *Multidimensional models of perception and cognition* (pp. 229–252). Hillsdale, NJ: Erlbaum.

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. (1997). Recognition memory ROCs for item and associative information: The contribution of recollection and familiarity. *Memory & Cognition, 25,* 747–763.

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

Yonelinas, A. P. (1999b). Recognition memory ROCs and the dual-process signal detection model: Comment on Glanzer, Kim, Hilford, and Adams. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 514–521.

Yonelinas, A. P. (2001). Consciousness, control, and confidence: The 3 Cs of recognition memory. *Journal of Experimental Psychology: General, 130,* 361–379.

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., Dobbins, I., Szymanski, M. D., Dhaliwal, H. S., & King, L. (1996). Signal detection, threshold, and dual process models of recognition memory: ROCs and conscious recollection. *Consciousness and Cognition, 5,* 418–441.

Yonelinas, A. P., Kroll, N. E. A., Dobbins, I. G., Lazzara, M., & Knight, R. T. (1998). Recollection and familiarity deficits in amnesia: Convergence of remember/know, process dissociation, and receiver operating characteristic data. *Neuropsychology, 12,* 1–17.

Yonelinas, A. P., Kroll, N. E. A., Dobbins, I. G., & Soltani, M. (1999). Recognition memory for faces: When familiarity supports associative recognition judgments. *Psychonomic Bulletin and Review, 6,* 654–661.

Yonelinas, A. P., Kroll, N. E. A., Quamme, J. R., Lazzara, M. M., Sauve, M. J., Widaman, K. F., & Knight, R. T. (2002). Effects of extensive temporal lobe damage or mild hypoxia on recollection and familiarity. *Nature Neuroscience, 5,* 1236–1241.

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.

Received December 16, 2005
Revision received April 3, 2006
Accepted May 6, 2006 ■