



Published in final edited form as:

Hippocampus. 2010 November ; 20(11): 1195–1205. doi:10.1002/hipo.20854.

Measuring Recollection and Familiarity in the Medial Temporal Lobe

John T. Wixted^{1,*}, Laura Mickes¹, and Larry R. Squire^{1,2,3,4}

¹ Department of Psychology, University of California, San Diego, La Jolla, California

² Department of Psychiatry, University of California, San Diego, La Jolla, California

³ Department of Neurosciences, University of California, San Diego, La Jolla, California

⁴ Veterans Affairs Medical Center, San Diego, California

Abstract

Many recent studies have investigated how the structures of the medial temporal lobe (MTL) support recollection and familiarity, which are two processes widely thought to support recognition memory. The behavioral methods that are used to isolate recollection and familiarity in neuroimaging and lesion studies typically assume that recollection is a categorical process and not a continuous process. A categorical process is one that either occurs or does not occur for a particular test item (yielding high confidence and high accuracy when it does occur), whereas a continuous process is one that comes in degrees (yielding varying degrees of confidence and accuracy). Studies suggesting that the hippocampus selectively supports the recollection process (such as those that use the Remember/Know procedure or rely on Receiver Operating Characteristic analysis) generally depend on the categorical view of recollection, but much recent evidence suggests that recollection is a continuous process. If recollection is a continuous process (i.e., if recollection comes in degrees), then evidence that has been taken to mean that the hippocampus selectively supports recollection is also compatible with the idea that the hippocampus supports both recollection and familiarity. We suggest that an alternative method can be used to effectively investigate recollection and familiarity in the MTL, one that is valid whether recollection is a categorical or a continuous process.

Keywords

hippocampus; recollection; familiarity; remember/know procedure; ROC analysis

INTRODUCTION

Recollection and familiarity are two memory processes that are widely thought to underlie the ability to recognize an item as having been previously encountered (Atkinson and Juola, 1974; Mandler, 1980; Jacoby, 1991; Yonelinas, 2002). Recollection involves the retrieval of specific details associated with a test item (e.g., remembering that the word “atlas” was presented in the top left quadrant of the screen), whereas familiarity involves simply knowing that an item was previously encountered even though nothing specific about its prior occurrence can be recalled. Many recent studies have investigated how the structures

*Correspondence to: John T. Wixted, Department of Psychology, University of California, San Diego, La Jolla, CA 92093. jwixted@ucsd.edu.

Additional Supporting Information may be found in the online version of this article.

of the medial temporal lobe (MTL) support these two memory processes. To do so, they have relied on various behavioral methods to analyze recognition decisions in terms of recollection and familiarity. The behavioral methods most commonly used for this purpose involve the analysis of confidence ratings, often made using a six-point scale, and the use of the Remember/Know procedure, in which participants are asked to indicate directly for each recognition decision they make whether the decision was based on recollection (by choosing “Remember”) or familiarity (by choosing “Know”). These procedures have been used in an effort to isolate recollection-based and familiarity-based decisions in neuroimaging experiments (e.g., Wheeler and Buckner, 2004), and they have also been used to measure the integrity of the recollection and familiarity processes in patients with MTL lesions (e.g., Aggleton et al., 2005).

Although a reasonably consistent pattern of empirical results has been observed in studies that use these procedures, the interpretation of those results turns out to hinge on whether recollection is construed as a categorical process or as a continuous process. This point has not been widely appreciated, and the reason why this issue is of critical importance is the main focus of this article.

Continuous Versus Categorical Memory Processes

There is broad agreement that familiarity is a continuous process. That is, individual items on a recognition memory test can be associated with low, medium, or high degrees of familiarity. As such, decisions based on familiarity can be associated with a range of confidence and a range of accuracy. Whether or not the recollection process shares this property with familiarity is a matter of some dispute.

By some accounts, recollection is a *categorical* process with respect to confidence and accuracy (e.g., Yonelinas, 1994). According to this idea, recollection succeeds for some items (in which case, the recognition decision is based solely on recollection) and fails for other items (in which case, the recognition decision is based solely on familiarity). When recollection succeeds, the recognition decision is assumed to be characterized by high confidence and high accuracy. When recollection fails, the confidence and accuracy of the decision are assumed to depend on the strength of the familiarity signal (with higher familiarity being associated with higher confidence and accuracy).

According to a different view, recollection—like familiarity—is a continuous process with respect to both confidence and accuracy (Wixted, 2007). That is, individual test items can be associated with low, medium, or high degrees of recollection and correspondingly low, medium, or high degrees of confidence and accuracy. From this perspective, the fundamental difference between recollection and familiarity is not that one process is categorical and the other continuous; instead, it is that one process involves the retrieval of source information and the other does not. The recollection signal, whether weak or strong, contains source information. The stronger the recollection signal, the higher the confidence and accuracy it supports. By contrast, the familiarity signal, whether weak or strong, is largely devoid of source detail. Even in the relative absence of source information, the stronger the familiarity signal, the higher the confidence and accuracy it supports.

Why does it matter whether recollection is construed as a categorical or a continuous process? The reason is that a large body of evidence, which has been interpreted to mean that the hippocampus selectively supports the recollection process, presumes that recollection is a categorical process [see Eichenbaum et al. (2007) for a review of such evidence]. Much of this evidence is based on estimates of recollection and familiarity obtained from fitting the Dual Process Signal Detection (DPSD) model (Yonelinas, 1994) to confidence-based Receiver Operating Characteristic (ROC) data and from the use of the

Remember/Know procedure (Tulving, 1985) to interpret neuroimaging data in terms of recollection and familiarity. Yet, as we explain in subsequent sections, if recollection is a continuous process, then this evidence cannot be used in support of the idea that the hippocampus selectively subserves recollection. In fact, the same evidence would be compatible with the idea that the hippocampus subserves both recollection and familiarity.

One potentially confusing aspect of the debate over the nature of recollection is the recent suggestion that the DPSD model and the Remember/Know procedure do not actually assume that recollection is a categorical process but instead allow for the possibility that recollection is a continuous process. Parks and Yonelinas (2007, 2009), for example, suggested that the DPSD model merely assumes that there is some level of memory strength (the “threshold” level of strength) below which recollection fails. Above that threshold level, recollection is assumed to be continuous in the sense that one may recollect a little or a lot about the study episode. Thus, according to this argument, the only difference between the two views is that one view assumes that memory strength can be low enough such that recollection fails (the threshold view), whereas the other view assumes that no matter how weak memory is, some degree of recollection is always involved. However, this is not the important point of contention between the two views.

The key point of contention involves an assumption made by the DPSD model about the relationship between the occurrence of any degree of recollection and the confidence that is expressed in a recognition decision. Confidence ratings are usually made using a six-point scale (1 = Sure New, 2 = Probably New, 3 = Maybe New, 4 = Maybe Old, 5 = Probably Old, and 6 = Sure Old). According to the DPSD model, whenever any degree of recollection occurs, the recognition decision is made with high confidence (i.e., six) and high accuracy (Yonelinas, 1994). If this assumption is correct, it follows that all other levels of confidence (one through five) reflect familiarity-based decisions. By contrast, the continuous recollection account holds that recollection is low for ratings of four (Maybe Old), higher for ratings of five (Probably Old), and higher still for ratings of six (Sure Old). The interpretation of ROC data in terms of the DPSD model and neuroimaging data in terms of the Remember/Know procedure hinges on which view of recollection is correct.

Confidence Ratings and ROC Analysis

The assumption that any degree of recollection yields the highest level of confidence (and that all lower levels of confidence reflect familiarity-based decisions) provides the basis for the estimates of recollection and familiarity obtained from fit-ting the DPSD model to confidence-based ROC data. This is true of every experiment that has ever relied on that method. Thus, if one relies on ROC evidence interpreted in terms of the DPSD model to support the claim that the hippocampus selectively supports the recollection process, then one is simultaneously endorsing the claim that any degree of recollection yields the highest level of confidence (because that assumption is what allows the DPSD model to extract estimates of recollection and familiarity from ROC data). If this assumption about the relationship between recollection and confidence is not valid, then the interpretation of the ROC data in terms of the DPSD model is not valid. As we describe next, the evidence consistently shows that varying degrees of recollection are associated with varying levels of confidence and accuracy (not just with the highest level of confidence).

Perhaps the most straightforward way to investigate the nature of recollection is to measure the relationship between the confidence expressed in a source recollection decision and the accuracy of that decision. In a source-memory procedure, items are presented on a study list along with a to-be-recollected detail (e.g., the items are presented in different colors or in different locations on the screen). At test, the items are presented in neutral fashion (e.g., in black or at the center of the screen), and participants are asked to recollect the source detail

that was associated with the item at study and to express confidence in their source decision. The categorical recollection view predicts that the relationship between confidence and accuracy will be a discontinuous step function, with the highest level of confidence associated with accurate source recollection and all lower levels of confidence associated with chance accuracy. The continuous view, by contrast, predicts a graded relationship between source confidence and source accuracy. Recently, Mickes et al. (2009) used a source-memory procedure in which items were presented in two different colors (red or blue) or, in a second experiment, in two different locations on the screen (top or bottom). On the subsequent source-memory test, the items were presented in source-neutral fashion and the participants were asked to rate their confidence about both kinds of source information using a 20-point scale (1 = Sure Source A and 20 = Sure Source B, where A and B refer to either red and blue or top and bottom). As shown in Figure 1, the observed relationship was clearly continuous over the 20 levels of confidence. Other studies have reported similar findings (Slotnick and Dodson, 2005; Gold et al., 2006; Kirwan et al., 2008).

Another way that the nature of recollection can be addressed involves testing how confidence in an old/new decision (rather than confidence in a source memory decision) is related to the accuracy of a subsequent source recollection decision. On the recognition test, each item is first presented for an old/new confidence rating (usually using a scale ranging from 1 = Sure New to 6 = Sure Old), and then this rating is followed by a test of source recollection (e.g., “in which quadrant of the screen was the word presented?”). If recollection is a categorical process that yields high confidence whenever it occurs, then high-confidence old decisions (e.g., ratings of six on a six-point scale) should be associated with accurate source recollection, but old/new decisions made with lower levels of confidence (e.g., ratings of one through five) should be associated with chance accuracy on the source-memory test. That is, once again, source accuracy plotted as a function of old/new confidence should abruptly increase as a discontinuous step-function when confidence reaches the highest level of six. This is the relationship that the DPSD model assumes to hold whenever it is used to extract estimates of recollection and familiarity from ROC data.

Yonelinas (2001) reported findings that showed a step-function relationship between old/new confidence and source accuracy, as predicted by a categorical account of recollection. Had this finding been reinforced by subsequent research, it would have served to validate the key assumption made by the DPSD model. However, studies conducted since that time have consistently found that the relationship is graded. Typically, when a six-point scale is used, ratings of four (Maybe Old) are associated with low but above-chance source accuracy, ratings of five (Probably Old) are associated with higher levels of source accuracy, and ratings of six (Sure Old) are associated with still higher levels of source accuracy (e.g., Slotnick et al., 2000; Slotnick and Dodson, 2005; Gold et al., 2006). Thus, the weight of evidence based on this procedure also supports the continuous view of recollection. What this means is that confidence-based ROC data that have been interpreted to mean that the hippocampus selectively subserves recollection based on parameter estimates obtained from the DPSD model do not support that conclusion (because the parameter estimates are based on an assumption about recollection that is not true).

Findings with rats involving the ROC procedure have also been advanced in support of the idea that the hippocampus selectively subserves the recollection process (Fortin et al., 2004; Sauvage et al., 2008). These ROCs were obtained using a biasing manipulation, and they were interpreted using the DPSD model—the same model that has often been used to interpret confidence-based ROCs in humans. Interpreted in terms of that model, the rat data are consistent with the idea that the hippocampus plays a role in recollection but does not play any role in familiarity. However, this interpretation is dependent on the assumption that recollection in the rat is a categorical process. If it is instead a continuous process (as it is in

humans), then the results of those studies would have a different meaning. For example, as noted elsewhere (Wixted and Squire, 2009), the ROC patterns they reported might result from any of a variety of as-yet-untested motivational and/or procedural variables associated with the unique biasing procedure that has been developed to obtain ROC data in rats (Fortin et al., 2005; Sauvage et al., 2008).

The main point is that the interpretation of ROC data in terms of the DPSD model—an interpretation that sometimes suggests that the hippocampus plays no role in familiarity—is fully dependent on the assumption that recollection is a categorical process. This is an important consideration in light of a growing body of evidence suggesting that recollection is instead a continuous process (e.g., Rutishauser et al., 2006, 2008; Kurilla and Westerman, 2010; Onyper et al., 2010; Slotnick, 2010). If recollection is a continuous process, then the estimates of recollection and familiarity obtained from fitting the DPSD model to ROC data cannot be used to support the idea that the hippocampus selectively subserves recollection. The problem is that such estimates, when derived from confidence-based ROCs in humans, tend to incorrectly suggest that weak memory (typical of patients with hippocampal lesions) reflects performance based purely on familiarity. If recollection is continuous, then weak memory would instead reflect performance based on lesser degrees of both recollection and familiarity. The only way to differentiate between these two interpretations is to equate for overall memory strength and then ask whether the shape of the ROC still differs for patients and controls (as should be the case if recollection is selectively impaired in the patients). Wais et al. (2006) equated for memory strength by using shorter lists for hippocampal patients (10 items) than for a group of matched controls (50 items). Once the memory strength was equated, ROC differences between the patients and controls were no longer evident. This result is consistent with the idea that the hippocampus subserves both recollection and familiarity. Similar considerations apply to the widely used Remember/ Know procedure.

The Remember/ Know Procedure

In the Remember/ Know procedure, participants are asked to directly indicate whether their recognition decisions are based on recollection or familiarity by saying “Remember” or “Know,” respectively. In studies that have used this procedure to investigate the role of the hippocampus in recollection and familiarity, recollection is assumed to be categorical with respect to Remember and Know judgments. That is, Remember judgments are assumed to reflect recollection, and Know judgments are assumed to reflect familiarity (which simply means that participants are assumed to be able to do what they are asked to do). By contrast, we assume that recollection is continuous with respect to Remember/ Know judgments in the same way that recollection is continuous with respect to confidence ratings. That is, Remember judgments reflect a high degree of recollection, whereas Know judgments reflect a lower degree of recollection. More generally, Remember judgments reflect strong memory (strong in both recollection and familiarity), whereas Know judgments reflect weaker memory (weaker in both recollection and familiarity). To illustrate why this strength confound matters, we consider how the Remember/ Know procedure has been used with fMRI to investigate the neuroanatomical basis of recollection and familiarity.

In one representative study, Eldridge et al. (2000) found that activity at retrieval was elevated in the hippocampus for Remember judgments relative to activity associated with forgotten items but was not elevated for Know judgments. If recollection is a categorical process, and if Remember judgments identify items for which recollection occurred, whereas Know judgments identify familiar items for which recollection did not occur, then this result would count as evidence that the hippocampus is associated selectively with the recollection process.

The absence of elevated activity associated with Know judgments relative to forgotten items is particularly interesting, because Know judgments are generally thought to be based on a strong sense of familiarity (and to therefore involve high confidence and high accuracy). Indeed, standard instructions stipulate that Know judgments should only be made when confidence is high. Consider, for example, the relevant portion of the widely used instructions for Know judgments from Rajaram (1993):

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

If Know judgments reflect an experience like this, then the failure to find any trace of elevated activity in the hippocampus associated with Know judgments would constitute evidence that the hippocampus plays no role in familiarity. Moreover, if recollection is a categorical process, it seems reasonable to suppose that Know judgments do in fact reflect strong, familiarity-based recognition (just as the instructions stipulate).

How should results like these be interpreted if recollection is instead a continuous process? In that case, Remember judgments would not represent the occurrence of recollection, *per se*, but would instead indicate that *enough* recollection had occurred for the participant to declare that the item was Remembered. This view of recollection does not substantially alter the interpretation of Remember judgments because, whether recollection is categorical or continuous, Remember judgments are assumed to identify decisions involving a considerable degree of recollection. However, the continuous view of recollection does substantially alter the interpretation of Know judgments. If recollection is a continuous process, then Know judgments do not indicate the absence of recollection and the presence of a strong sense of familiarity. Instead, they indicate decisions for which the recollection signal is not strong enough for the item to be judged Remembered. In other words, Remember judgments indicate relatively strong recollection, whereas Know judgments indicate a lesser degree of recollection. This is a key point that follows from the continuous view of recollection.

The continuous view of recollection not only holds that Know judgments involve some degree of recollection (instead of the absence of recollection); it also raises the possibility that Know judgments reflect weak memory overall (including weak familiarity). If items that are well encoded tend to be strong in both recollection and familiarity, whereas items that are poorly encoded tend to be weaker in both recollection and familiarity, then items that are relatively strong in recollection (and, as a result, receive Remember judgments) are also likely to be relatively strong in familiarity. Similarly, items that are relatively weak in recollection (and, as a result, receive Know judgments) are also likely to be relatively weak in familiarity. If so, Know judgments would reflect relatively weak memory (weak in both recollection and familiarity) and would therefore be associated with lower confidence and lower accuracy than Remember judgments. This is a second key point that follows from the continuous view of recollection (i.e., Know judgments reflect relatively weak memory). Next, we briefly review the evidence, suggesting that Know judgments indeed do involve lesser degrees of recollection than Remember judgments and that they reflect weaker memory overall compared to Remember judgments.

Know judgments involve recollection—One approach to investigating whether Know judgments involve some degree of recollection is to test the relationship between Remember/Know judgments and source accuracy. This method is similar to the procedure described earlier to test whether confidence ratings are associated with varying degrees of recollection. The categorical view of recollection predicts that Remember judgments will be associated with accurate source recollection, whereas Know judgments will be associated

with source recollection accuracy that is no better than chance (because, on this view, Know judgments identify familiarity-based decisions that are devoid of recollection). By contrast, the continuous view of recollection holds that Remember judgments will be associated with a higher level of source recollection accuracy than Know judgments, but both will exceed chance.

Almost invariably, the pattern of results predicted by the continuous recollection account has been observed (e.g., Eldridge et al., 2005). Remember judgments are associated with better source accuracy than Know judgments, and both kinds of judgment are associated with above-chance accuracy. Wais et al. (2008) reported the same pattern, and they tested a possible alternative explanation that could reconcile the categorical recollection account with findings like these. Specifically, because Know judgments are typically made before source recollection is tested, it is possible that the initial Know judgment (for items declared old) is based purely on familiarity as the categorical recollection account assumes but that a second retrieval attempt prompted by the source recollection question is occasionally successful (thereby accounting for above-chance source recollection associated with Know judgments). Contrary to this hypothesis, Wais et al. (2008) found that the same pattern was observed (namely, an intermediate level of above-chance source recollection accuracy associated with Know judgments) even when source recollection was tested before participants were asked to make Remember/Know judgments for the items declared old.

A recent neuroimaging study (Johnson et al., 2009) offers further support for the continuous recollection view. Their study was based on the idea that recollection (but not familiarity) involves the reinstatement of the distributed pattern of cortical activity that occurred when an episode was initially encoded. To investigate how reinstatement is related to Remember/ Know judgments, they used multivoxel pattern-classification analyses of fMRI data and found that cortical reinstatement was evident for both Remember and Know judgments (though to a lesser degree for Know judgments compared to Remember judgments). From this finding, they concluded that the reinstatement of encoding activity, per se, is not sufficient to elicit a Remember judgment. Instead, the reinstatement must exceed a certain criterion level (i.e., recollection must be strong *enough*) for a participant to declare that an item is Remembered. They interpreted these results to support the continuous view of recollection over the categorical view.

Know judgments reflect relatively weak memory—Generally speaking, memories are said to be strong when they are associated with high confidence, high accuracy, and fast reaction times. If recollection is a continuous process, as most of the evidence just reviewed suggests is the case, then Remember judgments likely reflect strong memories (strong in both recollection and familiarity), whereas Know judgments reflect weaker memory (weaker in both recollection and familiarity). In agreement with this view, Remember judgments are known to be made with higher confidence and higher accuracy than Know judgments (Tulving, 1985; Dunn, 2004; Wixted and Stretch, 2004). Indeed, in Eldridge et al. (2000), accuracy for Remember judgments was 0.97 (Remember hit rate = 0.42 and Remember false alarm rate = 0.01), whereas accuracy for Know judgments was 0.48, which is no better than chance (Know hit rate = 0.31 and Know false alarm rate = 0.33).ⁱ In light of this information, the fact that elevated activity was not detected for Know judgments is perhaps not surprising. In addition, much evidence has accumulated showing that old/new decisions associated with Remember judgments are made more quickly than old/new decisions associated with Know judgments (e.g., Dewhurst et al., 2006). The cumulative body of

ⁱThe overall hit rate, the Remember hit rate, and the Remember and Know false alarm rates were reported by Eldridge et al. (2000). The Know hit rate (.31) was computed by subtracting the Remember hit rate (.42) from the overall hit rate (.73). Accuracy (proportion correct) is equal to the hit rate for a particular judgment divided by the hit rate plus the false alarm rate for that judgment.

evidence has established beyond any reasonable doubt that Know judgments reflect weaker memory, on average, than Remember judgments.

Might Remember/Know judgments, as typically used, be useful for interpreting neuroimaging data despite the strength confound? After all, even if Know judgments involve some degree of recollection and Remember judgments involve some degree of familiarity (i.e., even if Remember/Know judgments are not “process pure”), Remember judgments surely involve considerably *more* recollection than Know judgments. That being the case, if elevated hippocampal activity is identified by a Remember-Know contrast, it would seem to follow that the hippocampus supports recollection. However, the problem is not just that Know judgments contain some degree of recollection and that Remember judgments contain some degree of familiarity. Instead, the problem is that Remember judgments may reflect higher degrees of *both* recollection and familiarity (i.e., they reflect stronger memories) than Know judgments. Thus, one could just as easily argue that elevated hippocampal activity identified by a Remember-Know contrast suggests that the hippocampus supports familiarity. This is why evidence based on the Remember/Know procedure—as it is typically used—is inconclusive with respect to the role of the hippocampus in recollection and familiarity.

Confidence Ratings and Remember/Know Judgments in a Single Scale

In a common variant of the Remember/Know procedure, participants are asked to rate test items using a continuous scale of confidence that culminates in a Remember judgment. For example, for each item on the recognition test, the participant might be asked to use a scale such as 1–2–3–4–R, where 1 = Sure New, 2 = Maybe New, 3 = Maybe Old, 4 = Sure Old, and R = Remember (e.g., Yonelinas et al., 2005). If recollection is a categorical process, and if Remember judgments identify decisions based on the occurrence of recollection, then confidence ratings of one through four reflect decisions based on the continuous familiarity process (analogous to Know judgments). In effect, this method is the procedural instantiation of the DPSD model, which holds that recollection is a categorical process with respect to confidence, whereas familiarity is a continuous process with respect to confidence. According to the DPSD model, any degree of recollection yields the highest level of confidence. Thus, the scale need not allow for different levels of confidence for Remember judgments as it does for judgments thought to be based on familiarity.

The results of several fMRI studies using a rating scale like this one have found that activity in the hippocampus does not increase as confidence increases from one to four but does abruptly increase in discontinuous fashion for ratings of R (e.g., Yonelinas et al., 2005; Montaldi et al., 2006; Cohn et al., 2009). If recollection is a categorical process, then results like these again suggest that the hippocampus does not play a role in familiarity-based decisions but does play a role in recollection-based decisions. However, if recollection is a continuous process, then the results would have a different interpretation. In particular, ratings of three and four (Maybe Old and Sure Old, respectively) would not indicate “old” decisions based only on familiarity but would instead indicate “old” decisions that involve increasing degrees of recollection (as well as familiarity). A rating of R would indicate a still greater degree of recollection—one that happens to be high enough for the participant to choose the R option. In addition, a rating of R would likely indicate stronger memory overall (i.e., stronger in both recollection *and* familiarity, and, as a result, higher in old/new accuracy compared to memories associated with ratings of four or less).

Consider, for example, a study reported by Yonelinas et al. (2005). They used a rating scale of 1–2–3–4–R, where 1 = Sure New, 2 = Unsure New, 3 = Unsure Old, 4 = Sure Old, and R = Recollected. To check whether a strength confound exists, one can simply compute accuracy for “old” decisions associated with ratings of three, four, and R. As shown in Table

1, in Yonelinas et al. (2005), the accuracy scores for three, four, and R (computed from the hit and false alarm rate data presented in their Table 1) were 0.49, 0.82, and 0.95, respectively. That is, ratings of three (Unsure Old) reflected chance performance, ratings of four (Sure Old) reflected moderately accurate performance, and ratings of R (recollected) reflected extremely accurate performance. Thus, R responses reflected stronger memories than responses of three or four. The fact that elevated activity in the hippocampus was selectively observed for ratings of R was taken to mean that the hippocampus selectively subserves recollection, but, in light of those accuracy scores, the results may instead mean that activity in the hippocampus is detectable when memory is strong (and accurate) but not when it is weaker (and less accurate). This is true even though the weaker memories associated with ratings of four likely involved considerable recollection (but not enough recollection for the participant to choose a rating of R).

In a similar study, Montaldi et al. (2006) used a rating scale of F1, F2, F3, and R for “old” decisions and found that hippocampal activity was selectively elevated for R judgments (the typical result). However, as shown in Table 1, old/new accuracy scores (computed from data reported in their Tables 2 and 3, as described in the Supporting Information) were 0.54, 0.68, 0.82, and 0.89 for F1, F2, F3, and R, respectively. Thus, their results show that elevated hippocampal activity was detected when memory was strong (R), but not when memory was weaker (F3), even though considerable recollection may have been involved for ratings of F3 and perhaps even F2 (if recollection is a continuous process). Indeed, because recollection is usually associated with lower (as well as higher) levels of confidence, this scale could, in our view, be conceptualized as F1–F2–R3–R4 (with the higher confidence levels of R3 and R4 also being associated with increasing levels of familiarity). Viewed from this perspective, the fact that elevated activity was evident only for ratings of what could have been labeled R4 indicates that memory must be strong (whether that memory involves recollection or familiarity) for such activity to be detected in the hippocampus.

Using the data from Montaldi et al. (2006), Montaldi and Mayes (this issue) report new old/new accuracy scores that were computed at the level of the individual subject and then averaged [instead of computing the accuracy scores based on the mean hit and false alarm rate data presented in Tables 2 and 3 of Montaldi et al. (2006), as we did]. Computed at the level of the individual subject, the accuracy scores for F1, F2, F3, and R are reported to be 0.54, 0.65, 0.88, and 0.89, respectively. Because the F3 accuracy score increased considerably, whereas the other scores (including the accuracy score for R) remained about the same, their new calculation suggests that old/new accuracy was almost equal for F3 and R.

It is not clear why the F3 accuracy score is sensitive to how it is computed (more so than the other scores). It is also not clear why the accuracy score for R judgments is lower than the accuracy scores typically obtained for Remember judgments. For example, in Eldridge et al. (2000), Remember judgments were associated with an old/new accuracy score of 0.98, and the old/new accuracy score for Remember judgments in Yonelinas et al. (2005) was 0.95 (as shown in Table 1). One possibility is that an outlier in Montaldi et al. (2006) used the R option very inaccurately, which could yield an unrepresentatively low-average accuracy score (creating the false impression that memory strength for R judgments was lower than it actually was). Still, this new information suggests that there is ambiguity about whether strength was equated in Montaldi et al. (2006).

Even if Montaldi et al. (2006) did equate old/new memory strength for ratings of F3 and R, they did not equate for strength at the high level ordinarily associated with R judgments. That is, F3 accuracy was, at most, 0.88, whereas accuracy for R judgments is typically

closer to 0.95. If strength was in fact equated at this relatively low level, their results may indicate that hippocampal activity becomes detectable at a lower level of strength for recollection-based memories than it does for familiarity-based memories. This would be an interesting finding, but it would not indicate whether elevated hippocampal activity can be observed when familiarity-based memories are as strong as the recollection-based memories that are typically associated with Remember judgments.

Cohn et al. (2009) designed a neuroimaging experiment that attempted to eliminate the strength confound by matching strength for recollection-based and familiarity-based decisions at a high level of strength. They used a rating scale of 1–2–3–4–R, where 1 = Sure New, 2 = Unsure New, 3 = Unsure Familiar, 4 = Sure Familiar, and R = Recollected. Their instructions stressed that recollected items and highly familiar items (i.e., items rated four on the confidence scale) should differ only in terms of the presence or absence of recollection but not in terms of confidence or strength. That is, instructions were used in an effort to eliminate any strength confound that might exist and to ensure that memory was strong for ratings of four. In addition, during a practice session, participants were asked to justify their responses to ensure that they understood the instructions. These instructions did serve to minimize R false alarms. However, the old/new accuracy scores (computed from data presented in their Table 2) indicate that the instructions did not succeed in eliminating the strength confound. As shown in Table 1, old/new accuracy for ratings of three, four, and R was 0.42, 0.79, and 0.97, respectively (these figures were averaged over the “cued target” and “uncued target” conditions, which exhibited similar trends). Thus, despite instructions to use ratings of four and R when memory was equally strong but differed only in recollective content, ratings of four were associated with much weaker memory (as indicated by old/new accuracy) than ratings of R. Moreover, had recollection accuracy been independently assessed, our assumption is that ratings of four would also have been associated with considerable levels of recollection (not with the absence of recollection). Ratings of 4 would be devoid of recollection only if recollection is a categorical process.

To illustrate the problem associated with rating scales such as these, we asked 40 participants to study a list of 128 words, each presented for 2 s. Half of the words were presented at the top of the screen in either red or blue, and half were presented at the bottom of the screen in either red or blue. On the subsequent recognition test, the 128 targets were randomly intermixed with 128 foils, and each item was presented at the center of the screen (in black) for an old/new decision using a 1–2–3–4–R rating scale. Following this recognition test, the target items were rerandomized and presented again for a test of source recollection. That is, for each item, the participant was asked if the word was studied at the top or the bottom of the screen and whether it had been presented in red or blue. The results are presented in Figure 2.

With regard to old/new accuracy, the typical strength confound is evident in the fact that accuracy was lower for ratings of four than for ratings of R (and accuracy was lower still for ratings of three). In addition, source accuracy increased in continuous fashion across ratings of three, four, and R. Accuracy was above chance in each case, significantly so for ratings of four ($P < 0.01$) and R ($P < 0.01$). Furthermore, source accuracy for ratings of R was significantly higher than source accuracy for ratings of four ($P < 0.05$). This pattern is consistent with the continuous view of recollection and is also consistent with similar studies that have used either confidence ratings (e.g., Slotnick and Dodson, 2005) or Remember/ Know judgments (Wais et al., 2008) to test for the presence of recollection.

The findings shown in Figure 2 suggest that R judgments do not indicate the categorical occurrence of recollection. Instead, they indicate that enough recollection has occurred to make an R judgment. Moreover, the simplest explanation for why elevated activity in the

hippocampus is often observed for R judgments but not for ratings of four is that activity in the hippocampus is more likely to be detected when memory is strong than when it is weak (which is why activity is not usually elevated for ratings of four despite the presence of a considerable degree of recollection). Such a result could simply reflect nonlinear measurement scale properties of the BOLD signal within the hippocampus (Squire et al., 2007), or it could mean that hippocampal activity is typically associated with a relatively strong memory (Wais, 2008). Either way, if recollection is continuous, the important point is that the typical fMRI pattern obtained using the Remember/Know procedure would not indicate that the hippocampus selectively subserves recollection.

Findings like these do not necessarily imply that people cannot report valid information about the role of recollection and familiarity in their recognition decisions. Instead, the findings suggest that the usual procedures for eliciting such information place participants into a bind by asking them to characterize their experience of recollection as if it were a categorical process when it is instead a continuous process. What participants do to cope with the response options that experimenters provide to them is quite reasonable; namely, they choose “Remember” when recollection is strong enough and choose “Know” otherwise. Unfortunately, this reasonable strategy introduces the memory strength confound that plagues virtually every study that relies on the Remember/Know procedure [with Montaldi et al. (2006), being a possible exception].

Eliminating the strength confound (and also maintaining a high level of accuracy) to investigate whether the hippocampus subserves familiarity as well as recollection is not easy (so instructions alone may not suffice), but it may be possible. Later, we describe a possible way forward by proposing a method to eliminate the strength confound that should be compelling whether one views recollection as a categorical or a continuous process (thereby avoiding this aspect of the debate altogether). Before considering that approach, we address another concept that has been introduced in an effort to account for evidence that otherwise would appear to describe recollection as a continuous process. That concept is “unitized familiarity.”

Unitized Familiarity

Earlier in this article, we reviewed a growing and consistent body of evidence suggesting that recollection is a continuous process, one that is also associated with Know judgments (not just Remember judgments) and with varying degrees of confidence and accuracy. An alternative proposal is that the apparently successful source recollection associated with Know judgments (and with a range of confidence ratings) is not an indication of continuous recollection after all. Instead, it has been suggested that test items can sometimes be “unitized” with source information and in that case be based on familiarity. Accordingly, accurate source decisions can be made based on a familiarity signal that includes information about source details (Diana et al., 2008). For example, an item that is presented in the upper left quadrant of the screen might somehow become unitized with that location such that later, when the item is presented at the center of the screen for a source memory decision, it might reactivate the unitized representation. This process, in turn, would produce a familiarity signal that could support accurate source memory. The concept of a unitized familiarity signal has been advanced to account for what otherwise might be taken as evidence of a recollection signal in the perirhinal cortex rather than only in the hippocampus as had been proposed (e.g., Haskins et al., 2008).

The main evidence that has been advanced in favor of the concept of unitized familiarity (and the interpretation of brain activity in terms of that concept) consists of confidence-based ROC data interpreted in terms of the DPSD model. The ROC data have been obtained from source memory and associative recognition procedures, which are typically considered

to be recollection-based recognition tasks that involve little or no contribution from item familiarity (e.g., Yonelinas, 1997; Hockley and Consoli, 1999; Yonelinas, 1999). On a source-memory test, for example, if the items from the two sources (such as two different screen locations) are equally familiar, on average, then the familiarity signal cannot help one to decide whether the item was earlier presented at the top of the screen or the bottom of the screen. Recollection is usually considered to be necessary to answer a question like that. If recollection is a categorical process (such that any degree of recollection yields high confidence), then a confidence-based source ROC should be linear according to the DPSD model (Yonelinas, 1999). However, source ROCs are often curvilinear even though it is reasonable to suppose that the items from each source are equally familiar on average. The only way for the DPSD model to accommodate a curvilinear ROC is to assume that an item and its source have become unitized (and that a unitized trace yields a continuous familiarity signal).

Yet, a curvilinear source ROC merely indicates, in another way, that the relationship between confidence and accuracy on a source memory task is graded, as shown in Figure 1. Thus, a curvilinear ROC is as consistent with the idea that recollection is continuous as it is with a reinterpretation of source memory data in terms of unitized familiarity. Accordingly, other methods must be used to distinguish between these two possible views of a curvilinear source ROC.

To investigate whether a curvilinear ROC reflects continuous recollection or unitized familiarity, Mickes et al. (in press) used an associative recognition procedure with confidence ratings (1 = Sure Rearranged and 6 = Sure Intact) in which participants studied word pairs and were later asked to distinguish between intact and rearranged pairs. As in the case of source-memory tasks, item familiarity should offer no help to associative recognition tasks, because the items of intact and rearranged pairs are, on average, equally familiar. Indeed, no matter how familiar the two items of a test pair are, item familiarity cannot help the subject to decide whether the pair is intact or rearranged. Accordingly, it is often assumed that accurate performance is achieved by recollecting what words were associated at study.

Mickes et al. (in press) found that, for pairs that had been studied only once, the ROC was relatively linear (which the DPSD model interprets as a sign of categorical recollection). When pairs had been studied multiple times (i.e., when memory was stronger), the ROC became more curvilinear (which the DPSD model interprets as a sign of unitized familiarity). However, compared to the weakly remembered pairs, and contrary to the interpretation provided by the DPSD model, the strongly remembered pairs were associated with a much higher probability of Remember judgments (usually thought to be indicative of recollection, not unitized familiarity) and also with a much higher probability of successful cued recall on a subsequent memory test (a direct test of recollection). Thus, as intuition would suggest, strengthening memory increased recollection considerably, and it also increased the curvilinearity of the ROC. The implication is that recollection is a continuous process that yields curvilinear ROCs. Accordingly, it is problematic to rely on curvilinear ROC data to support the hypothetical construct of unitized familiarity. More to the point, showing that Condition A is associated with a curvilinear ROC and that Condition A is also associated with elevated activity in perirhinal cortex (e.g., Haskins et al., 2008) cannot be taken to suggest that perirhinal cortex subserves unitized familiarity, because that result is equally compatible with the idea that perirhinal cortex subserves continuous recollection.

This line of argument does not mean that items cannot ever be unitized (e.g., the words “black” and “berry” presented on a study list might be unitized into “blackberry,” which could then be recognized on a later test based on familiarity). Rather, we suggest that the

evidence that has been used to validate the notion that information has been unitized—namely, a curvilinear ROC in a source memory or associative recognition procedure—does not warrant that interpretation. If recollection is a continuous process, then recollection will also yield a curvilinear ROC.

The Use and Misuse of the Labels “Recollection” and “Familiarity”

These considerations raise an important point about how empirical results are presented in the research literature. Often, the column headings of tables that present data in research articles are labeled “recollection” and “familiarity” (e.g., Eichenbaum et al., 2007). The same holds true for figures that present patterns of brain activity (Dasselar et al., 2006). That is, some patterns are labeled “recollection” and others are labeled “familiarity.” Similarly, when illustrating hypothetical patterns of ROC data, one pattern is sometimes labeled “recollection” and the other “familiarity” (e.g., Fortin et al., 2004). We suggest that this practice be eliminated. At a minimum, it should be modified, so that readers can better appreciate the strong theoretical assumption about the nature of recollection that gives license to the use of such labels. In particular, such labels, if they are used at all, should be used with a disclaimer making it clear to readers that the labels apply if recollection is a categorical process but do not apply if recollection is associated with varying levels of confidence and accuracy (as much evidence suggests is the case).

A Way Forward?

The fact that the strength confound is ubiquitous in studies intending to distinguish recollection from familiarity does not mean that more effective methods cannot be developed to eliminate the confound. One way to address the issue might be to revisit the original intent of the Remember/Know procedure. As indicated earlier, Know judgments were originally intended to identify *strong* familiarity-based decisions (Rajaram, 1993). However, in practice, Know judgments identify weaker memories than Remember judgments that involve lesser degrees of recollection (and, perhaps, lesser degrees of familiarity as well). One possible solution would be to compare brain activity associated with Remember and Know judgments that have first been equated for both confidence and accuracy (and, perhaps, reaction time).

In this regard, Wixted and Mickes (in press) recently investigated the source-memory content of Remember and Know judgments that were equated for strength in terms of confidence and accuracy. After studying a list of words that were presented in one of two colors and in one of two screen locations, participants were presented with a recognition test in which each item was presented for an old/new confidence rating, followed by a Remember/Know/Guess judgment, followed by a source-recollection test (“Was this item presented in red or blue?”; “Was this item presented at the top or the bottom of the screen?”). The old/new confidence rating was made using a 20-point scale (1 = Sure New and 20 = Sure Old). As usual, Remember judgments were associated with higher confidence and higher accuracy (both for the old/new decisions and for the source decisions) than Know judgments (which were nevertheless associated with greater-than-chance source recollection). However, when only the high-confidence old decisions were analyzed (i.e., decisions associated with a confidence rating of 20), old/new accuracy was equated at ~91% correct for Remember and Know judgments. The interesting finding was that source-recollection accuracy differed considerably under these conditions. More specifically, Remember judgments were associated with much more accurate source recollection than Know judgments (80% correct vs. 65% correct). Thus, even when Remember and Know judgments are matched for memory strength, Remember judgments can reflect better memory for source content than Know judgments and in this way can differ qualitatively from Know judgments. Know judgments were still associated with above-chance source

recollection (as usual), but the level of source recollection was relatively low. Accordingly, it seems that the high-memory strength associated with confidence ratings of 20 as well as with Know judgments can be primarily attributed to strong familiarity. The point is that Remember and Know judgments were equated for memory strength (in terms of confidence and old/new accuracy). Nonetheless, Remember judgments were based largely on recollection, and Know judgments were based largely on familiarity. Given this circumstance, it would be useful to investigate hippocampal activity in association with these decisions (i.e., for equally strong memories that nevertheless differ in source memory content). If activity is selectively elevated for Remember judgments under these circumstances, this result would be consistent with the idea that the hippocampus plays a role in recollection but not familiarity. If activity is elevated for both Remember and Know judgments, this result would be consistent with the idea that the hippocampus plays a role in both recollection and familiarity. Despite many Remember/Know investigations to date, none have addressed this question.

CONCLUSION

From the arguments presented above (namely, that the methods used to investigate the neuroanatomical basis of recollection and familiarity suffer from a strength confound), it would be easy to conclude that the continuous view of recollection leads directly to the claim that the functional organization of the MTL should be understood in terms of memory strength. However, this is not the case. The key point is that the functional organization of the MTL will be best understood in terms unrelated to the distinction between recollection and familiarity, not that memory strength is an organizing principle for the MTL. Nor does it mean that any functional difference between the two structures should disappear once the strength has been equated. Squire et al. (2007) discussed alternative ideas about the functional organization of the MTL that may be more promising. Indeed, in our view, an overemphasis on the distinction between recollection and familiarity has drawn attention away from what may prove to be more productive lines of inquiry.

Acknowledgments

Grant sponsor: National Institute of Mental Health; Grant numbers: 082892 and 24600; Grant sponsor: Medical Research Service of the Department of Veterans Affairs.

References

- Aggleton JP, Vann SD, Denby C, Dix S, Mayes AR, Roberts N, Yonelinas AP. Sparing of the familiarity component of recognition memory in a patient with hippocampal pathology. *Neuropsychologia*. 2005; 43:1810–1823. [PubMed: 16154457]
- Atkinson, RC.; Juola, JF. Search and decision processes in recognition memory. In: Krantz, DH.; Atkinson, RC.; Suppes, P., editors. *Contemporary Developments in Mathematical Psychology*. San Francisco: Freeman; 1974. p. 243-290.
- Cohn M, Moscovitch M, Lahat A, McAndrews MP. Recollection versus strength as the primary determinant of hippocampal engagement at retrieval. *Proc Natl Acad Sci USA*. 2009; 106:22451–22455. [PubMed: 20007783]
- Daselaar SM, Fleck MS, Cabeza R. Triple dissociation in the medial temporal lobes: Recollection, familiarity, and novelty. *J Neurophys*. 2006; 96:1902–1911.
- Dewhurst SA, Holmes SJ, Brandt KR, Dean GM. Measuring the speed of the conscious components of recognition memory: Remembering is faster than knowing. *Consciousness Cogn*. 2006; 15:147–162.
- Diana RA, Yonelinas AP, Ranganath C. The effects of unitization on familiarity-based source memory: Testing a behavioral prediction derived from neuroimaging data. *J Exp Psychol Learn Mem Cogn*. 2008; 34:730–740. [PubMed: 18605864]

- Dunn JC. Remember-know: A matter of confidence. *Psychol Rev.* 2004; 111:524–542. [PubMed: 15065921]
- Eichenbaum H, Yonelinas AR, Ranganath C. The medial temporal lobe and recognition memory. *Annu Rev Neurosci.* 2007; 30:123–152. [PubMed: 17417939]
- Eldridge LL, Engel SA, Zeineh MM, Bookheimer SY, Knowlton BJ. A dissociation of encoding and retrieval processes in the human hippocampus. *J Neurosci.* 2005; 25:3280–3286. [PubMed: 15800182]
- Eldridge LL, Knowlton BJ, Furmanski CS, Bookheimer SY, Engel SA. Remembering episodes: A selective role for the hippocampus during retrieval. *Nat Neurosci.* 2000; 3:1149–1152. [PubMed: 11036273]
- Fortin NJ, Wright SP, Eichenbaum H. Recollection-like memory retrieval in rats is dependent on the hippocampus. *Nature.* 2004; 431:188–191. [PubMed: 15356631]
- Gold JJ, Smith CN, Bayley PJ, Shrager Y, Brewer JB, Stark CEL, Hopkins RO, Squire LR. Item memory, source memory, and the medial temporal lobe: Concordant findings from fMRI and memory-impaired patients. *Proc Natl Acad Sci USA.* 2006; 103:9351–9356. [PubMed: 16751272]
- Haskins AL, Yonelinas AP, Quamme JR, Ranganath C. Perirhinal Cortex supports encoding and familiarity-based recognition of novel associations. *Neuron.* 2008; 59:554–560. [PubMed: 18760692]
- Hockley WE, Consoli A. Familiarity and recollection in item and associative recognition. *Mem Cogn.* 1999; 27:657–664.
- Jacoby LL. A process dissociation framework: Separating automatic from intentional uses of memory. *J Mem Lang.* 1991; 30:513–541.
- Johnson JD, McDuff SGR, Rugg MD, Norman KA. Recollection, familiarity, and cortical reinstatement: A multivoxel pattern analysis. *Neuron.* 2009; 63:697–708. [PubMed: 19755111]
- Kirwan CB, Wixted JT, Squire LR. Activity in the medial temporal lobe predicts memory strength, whereas activity in the prefrontal cortex predicts recollection. *J Neurosci.* 2008; 28:10541–10548. [PubMed: 18923030]
- Kurilla BP, Westerman DL. Source memory for unidentified stimuli. *J Exp Psychol Learn Mem Cogn.* 2010; 36:398–410. [PubMed: 20192538]
- Mandler G. Recognizing: The judgment of previous occurrence. *Psychol Rev.* 1980; 87:252–271.
- Mickes L, Wais PE, Wixted JT. Recollection is a continuous process: Implications for dual-process theories of recognition memory. *Psychol Sci.* 2009; 20:509–515. [PubMed: 19320859]
- Mickes L, Johnson E, Wixted JT. Continuous recollection vs. unitized familiarity in associative recognition. *J Exp Psychol Learn Mem Cogn.* 2010; 36:843–863. [PubMed: 20565205]
- Montaldi D, Spencer TJ, Roberts N, Mayes AR. The neural system that mediates familiarity memory. *Hippocampus.* 2006; 16:504–520. [PubMed: 16634088]
- Onyper SV, Zhang YX, Howard MW. Some-or-none recollection: Evidence from item and source memory. *J Exp Psychol Gen.* 2010; 139:341–364. [PubMed: 20438255]
- Parks CM, Yonelinas AP. Moving beyond pure signal-detection models: Comment on Wixted. *Psychol Rev.* 2007; 114:188–202. [PubMed: 17227187]
- Parks CM, Yonelinas AP. Evidence for a memory threshold in second-choice recognition memory responses. *Proc Natl Acad Sci USA.* 2009; 106:11515–11519. [PubMed: 19564612]
- Rajaram S. Remembering and knowing: Two means of access to the personal past. *Mem Cogn.* 1993; 2:89–102.
- Rutishauser U, Mamelak AN, Schuman EN. Single-trial learning of novel stimuli by individual neurons of the human hippocampus-amygdala complex. *Neuron.* 2006; 49:805–813. [PubMed: 16543129]
- Rutishauser U, Schuman EM, Mamelak AN. Activity of human hippocampal and amygdala neurons during retrieval of declarative memories. *Proc Natl Acad Sci USA.* 2008; 105:329–34. [PubMed: 18162554]
- Sauvage M, Fortin NJ, Owens CB, Yonelinas AP, Eichenbaum H. Recognition memory: Opposite effects of hippocampal damage on recollection and familiarity. *Nat Neurosci.* 2008; 11:16–18. [PubMed: 18037884]

- Sauvage M, Beer Z, Eichenbaum H. Recognition memory: Adding a response deadline eliminates recollection but spares familiarity. *Learn Mem.* 2010; 17:104–108. [PubMed: 20154356]
- Slotnick SD. *Remember* source memory ROCs indicate recollection is a continuous process. *Memory.* 2010; 18:27–39. [PubMed: 19937493]
- Slotnick SD, Dodson CS. Support for a continuous (single-process) model of recognition memory and source memory. *Mem Cogn.* 2005; 33:151–170.
- Slotnick SD, Klein SA, Dodson CS, Shimamura AP. An analysis of signal detection and threshold models of source memory. *J Exp Psychol Learn Mem Cogn.* 2000; 26:1499–1517. [PubMed: 11185779]
- Squire LR, Wixted JT, Clark RE. Recognition memory and the medial temporal lobe: A new perspective. *Nat Rev Neurosci.* 2007; 8:872–883. [PubMed: 17948032]
- Tulving E. Memory and consciousness. *Can Psychol.* 1985; 26:1–12.
- Wais P. fMRI signals associated with memory strength in the medial temporal lobes: A meta-analysis. *Neuropsychologia.* 2008; 46:3185–3196. [PubMed: 18817791]
- Wais P, Wixted J, Hopkins R, Squire L. The hippocampus supports both the recollection and the familiarity components of recognition memory. *Neuron.* 2006; 49:459–466. [PubMed: 16446148]
- Wais PE, Mickes L, Wixted JT. Remember/know judgments probe degrees of recollection. *J Cogn Neurosci.* 2008; 20:400–405. [PubMed: 18004949]
- Wixted JT. Dual-process theory and signal-detection theory of recognition memory. *Psychol Rev.* 2007; 114:152–176. [PubMed: 17227185]
- Wixted JT, Mickes L. Memory strength vs. memory content: A signal-detection-based, dual-process interpretation of Remember/Know judgments. *Psychol Rev.* (in press).
- Wixted JT, Squire LR. Constructing receiver operating characteristics (ROCs) with experimental animals: Cautionary notes. *Learn Mem.* 2008; 15:687–690. [PubMed: 18772256]
- Wixted JT, Stretch V. In defense of the signal detection interpretation of Remember/Know judgments. *Psychon Bull Rev.* 2004; 11:616–641. [PubMed: 15581116]
- Yonelinas AP. Receiver-operating characteristics in recognition memory: Evidence for a dual-process model. *J Exp Psychol Learn Mem Cogn.* 1994; 20:1341–1354. [PubMed: 7983467]
- Yonelinas AP. Recognition memory ROCs for item and associative information: The contribution of recollection and familiarity. *Mem Cogn.* 1997; 25:747–763.
- Yonelinas AP. The contribution of recollection and familiarity to recognition and source memory: An analysis of receiver operating characteristics and a formal model. *J Exp Psychol Learn Mem Cogn.* 1999; 25:1415–1434. [PubMed: 10605829]
- Yonelinas AP. Components of episodic memory: The contribution of recollection and familiarity. *Philos Trans R Soc Ser B.* 2001; 356:1363–1374.
- Yonelinas AP. The nature of recollection and familiarity: A review of 30 years of research. *J Mem Lang.* 2002; 46:441–517.
- Yonelinas AP, Otten LJ, Shaw KN, Rugg MD. Separating the brain regions involved in recollection and familiarity in recognition memory. *J Neurosci.* 2005; 25:3002–3008. [PubMed: 15772360]

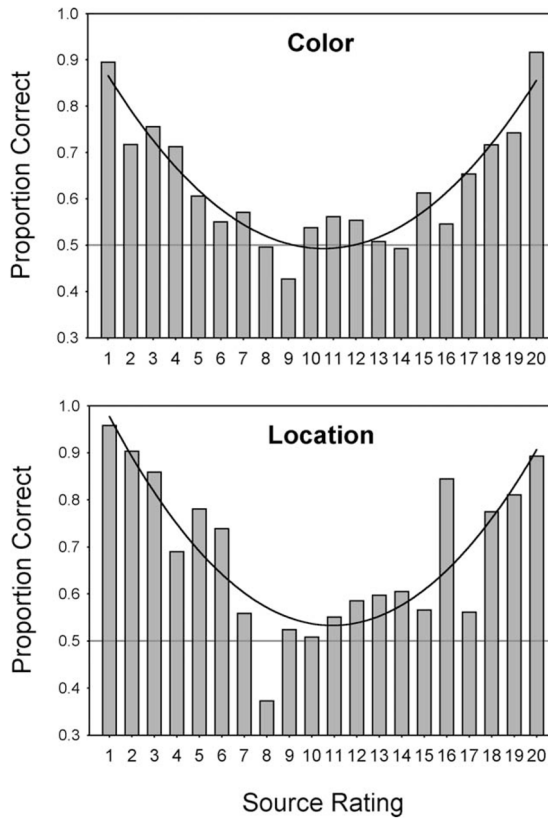


FIGURE 1.

Source recollection accuracy (proportion correct) as a function of confidence in the color and location versions of the source-memory task. On the 20-point confidence scale, 1 = Sure Red and 20 = Sure Blue for the color task, and 1 = Sure Top and 20 = Sure Bottom for the location task. The solid curve is a least squares fit of a second-order polynomial. Adapted from Mickes L, Wais PE, Wixted JT. 2009. Recollection is a continuous process: Implications for dual-process theories of recognition memory. *Psychol Sci* 20:509–515.

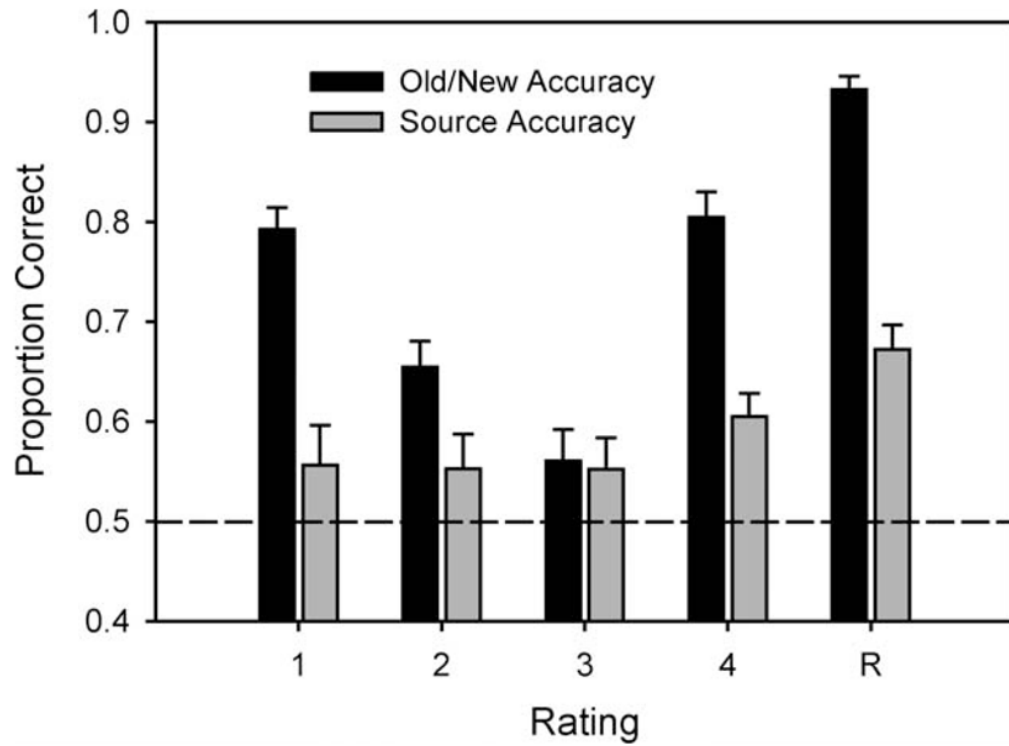


FIGURE 2. Old/new recognition accuracy and source recollection accuracy (proportion correct) for 40 subjects. The study list consisted of 128 words, and the source details were font color (red or blue) and screen location (top or bottom). Source accuracy scores were averaged over color and location because they showed identical trends.

TABLE 1

Old/New Accuracy Scores (Proportion Correct) for Old Decisions from Three Studies That Used a Continuous Confidence Scale for Familiarity-Based Decisions and an R Response for Recollection-Based Decisions

Study	Rating	Old/New Accuracy
Yonelinas et al. (2005)	3	0.49
	4	0.82
	R	0.95
Montaldi et al. (2006)	F1	0.54
	F2	0.68
	F3	0.82
	R	0.89
Cohn et al. (2009)	3	0.42
	4	0.79
	R	0.97

In Cohn et al. (2009) and Yonelinas et al. (2005), ratings of 3 and 4 represent low and high confidence for familiarity-based old decisions. In Montaldi et al. (2006), F1, F2 and F3 represent three levels of confidence in familiarity-based old decisions.