Journal of Experimental Psychology:
Learning, Memory, and Cognition
2014, Vol. 40, No. 5, 1205–1225

0278-7393/14/$12.00 http://dx.doi.org/10.1037/a0036846

Recognition Memory zROC Slopes for Items With Correct Versus Incorrect Source Decisions Discriminate the Dual Process and Unequal Variance Signal Detection Models

Jeffrey J. Starns and Caren M. Rotello
University of Massachusetts Amherst

Michael J. Hautus
University of Auckland

We tested the dual process and unequal variance signal detection models by jointly modeling recognition and source confidence ratings. The 2 approaches make unique predictions for the slope of the recognition memory zROC function for items with correct versus incorrect source decisions. The standard bivariate Gaussian version of the unequal variance model predicts little or no slope difference between the source-correct and source-incorrect functions. We also developed a "bounded" version of this model that did not permit below-chance source discrimination in any region of the evidence space. The bounded version predicts that the source-correct function should have a lower slope than the source-incorrect function. A bivariate version of the dual process signal detection model can predict slope differences in either direction, but it must predict a u-shaped source zROC function if the source-correct slope is lower than the source-incorrect slope. Across 4 experiments, results consistently showed that the recognition memory zROC function had a lower slope for items attributed to the correct source than items attributed to the incorrect source, and the source zROC function for words recognized with high confidence was linear. Only the bounded version of the unequal variance model successfully predicted the full pattern of results.

Keywords: bivariate signal detection model, recognition memory, source memory, unequal variance assumption, dual process theory

How do we know what happened to us in the past? At the most general level, we have cognitive mechanisms that provide evidence about the past and cognitive mechanisms that translate this evidence into explicit judgments about what we have and have not experienced. Here, we address a fundamental debate about how evidence informs memory decisions; namely, whether the same suite of systems collaborate to produce the evidence for all memory decisions or whether different decisions are based on evidence from separate, independent systems. These alternatives have inspired specific models that either base all decisions on a single, continuously varying strength value or base some decisions on continuous strength while basing other decisions on the threshold retrieval of qualitative information (Parks & Yonelinas, 2007; Wixted, 2007). Many attempts to discriminate these models have focused on fits to receiver operating characteristic (ROC) functions for either recognition memory or source memory tasks. We show that the accounts make distinct predictions for how selecting items with correct versus incorrect source judgments affects the recognition memory ROC function. As a result, joint modeling of ROC functions from both tasks provides a better test of the alternative accounts than modeling either task in isolation.

Receiver Operating Characteristic (ROC) Functions

An ROC function shows the relationship between the proportion of correct responses and the proportion of errors across a range of response biases (Egan, 1958; Macmillan & Creelman, 2005). Transforming the response proportions to *z*-scores yields the corresponding zROC function. ROC and zROC functions have played a prominent role in testing theories of recognition and source memory. In a standard recognition task, participants study a list of words and then take a test in which they are asked to respond "old" for words that were previously studied (i.e., targets) and "new" for words that were not (i.e., lures). The proportion of "old" responses to targets and lures defines the hit rate and false-alarm rate, respectively. ROC functions are usually formed by asking participants to rate their confidence that each item was studied. In a source task, participants learn items in a number of contexts or presentation formats ("sources"), and at test they must identify the source for each item. Again, ROC functions can be defined by obtaining confidence ratings from "Sure Source A" to "Sure Source B."

One prominent controversy in the ROC literature is whether memory judgments are based on a single, continuous strength value or on a mixture of a threshold recollection process and a continuous familiarity process (e.g., Dougal & Rotello, 2007;

This article was published Online First May 12, 2014.

Jeffrey J. Starns and Caren M. Rotello, Department of Psychology, University of Massachusetts Amherst; Michael J. Hautus, School of Psychology, University of Auckland.

Data collection was supported by National Institute of Mental Health Grant R01 MH60274 to Caren M. Rotello.

Correspondence concerning this article should be addressed to Jeffrey J. Starns, Department of Psychology, 441 Tobin Hall, University of Massachusetts Amherst, Amherst, MA 01003. E-mail: jstarns@psych.umass.edu

Dunn, 2004; Wixted & Squire, 2010; Yonelinas, 2002). In the dual process view, recollection describes the retrieval of qualitative information surrounding the learning event, whereas familiarity is a quantitative signal that tends to be stronger for recently encountered items (Yonelinas, 2002). The purely continuous view is adopted by theorists who propose that both recollection and familiarity are continuous variables, with decisions based on the combined strength across both types of information (Wixted, 2007; Wixted & Mickes, 2010). The continuous view is also consistent with models that compute memory evidence in terms of the degree of match between a memory probe and the contents of memory, including the global matching models (see Clark & Gronlund, 1996, for a review) and the newer Bayesian matching models (Dennis & Humphreys, 2001; McClelland & Chappell, 1998; Shiffrin & Steyvers, 1997).

Researchers have most commonly compared the continuous and dual process models by separately evaluating either item recognition or source memory ROC functions, but theorists have also developed models that fit both simultaneously (Banks, 2000; DeCarlo, 2003; Hautus, Macmillan, & Rotello, 2008; Klauer & Kellen, 2010; Onyper, Zhang, & Howard, 2010). In the current project, we explored a result that can only be defined by jointly considering item and source decisions. Specifically, we compared recognition memory zROC functions for items with correct versus incorrect source attributions using both previously published data and three new experiments. We applied a bivariate signal detection model similar to those developed by DeCarlo (2003) and Banks (2000). We also developed two new models: one that incorporates a threshold recollection process and one that places a boundary on the source evidence distributions to ensure that source discriminability does not reverse for low-strength items (similar to Model 2 proposed by Hautus et al., 2008). As we describe below, the models make distinct predictions for the difference between the source-correct and source-incorrect zROC slope. Although we do report quantitative fits, we will primarily evaluate the models by determining whether the qualitative patterns predicted by each model match the patterns that are consistently observed across experiments.

The Bivariate Gaussian (BG) Model

The Appendix provides the prediction equations for all of the models under consideration. Figure 1 displays the bivariate Gaussian (BG) model. The *x*-axis represents a source continuum with evidence typical of items from Source B on the left and evidence typical of items from Source A on the right. The *y*-axis is a continuum of item evidence, with higher values representing stronger evidence that the test item was studied. The sets of ovals show equal-density contours of the evidence distributions, with thicker lines representing higher probability densities. We made very similar distributional assumptions to those reported by DeCarlo (2003). Memory evidence for new items is represented by a zero-correlation bivariate Gaussian distribution with a mean of zero and a standard deviation of 1 in both dimensions. The mean item strength (μ_I) for words studied in both sources is higher than new words, and the variability in item strength (σ_I) is higher for studied items as well. This unequal-variance assumption accommodates the well-replicated finding that recognition memory zROC functions have a slope less than 1 (e.g., Glanzer, Kim,

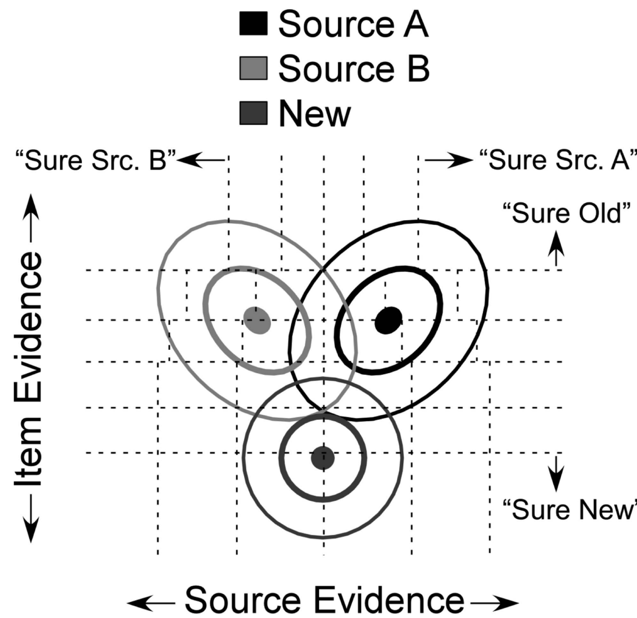

Figure 1. The bivariate Gaussian model of recognition and source confidence ratings. The sets of ovals represent bivariate Gaussian distributions for items studied in Source A (Src. A), items studied in Source B (Src. B), and items appearing for the first time at test (New). The solid dots show the means of each distribution, and the surrounding lines are equal-density contours with thicker lines for higher probability densities. Five criteria on the item dimension define regions associated with six item-confidence ratings. On the source dimension, the model has separate source criteria for "New" responses and for each of the three confidence levels of "Old" responses.

Hilford, & Adams, 1999; Ratcliff, Sheu, & Gronlund, 1992). The average source evidence (μ_S) for Source B items is farther to the left than Source A items, with new items in the middle (i.e., non-studied items are not associated with either source). Item and source evidence are correlated for the studied words, such that words with higher item strength also have more diagnostic source information. This property is represented by the "slant" of the distributions, and it is parameterized as a positive correlation (ρ) between item and source evidence for Source A items and a negative correlation for Source B items.

To map evidence states onto confidence responses, DeCarlo (2003) proposed a single set of linear decision bounds on both dimensions. We make the same assumption for criteria on the item strength dimension, and Figure 1 shows five criteria to segregate evidence values into responses on a 6-point confidence scale from "Sure New" to "Sure Old." However, we deviated from DeCarlo for the source criteria. Specifically, we allowed all of the models under consideration to establish different source confidence criteria for "new" responses and "old" responses made at each confidence level (Onyper et al., 2010). Figure 1 shows dispersed source confidence criteria for items called "new" (i.e., anything below the middle item recognition criterion), and converging criteria as items are recognized with higher confidence.

The converging criteria pattern displayed in Figure 1 has strong empirical support. Starns, Pazzaglia, Rotello, Hautus, and Macmillan (2013) demonstrated that participants are more willing to make high-confidence source judgments when they give higher ratings of item strength, even for items that were not studied in any source. Hautus et al. (2008) reported that a bivariate Gaussian model with likelihood ratio decision bounds outperformed a model with constant source criteria across different levels of item strength. One critical property of the likelihood bounds is that the source confidence criteria converge as item strength increases, which implements a strategy of using more high confidence source responses for stronger item evidence. Onyper et al. (2010) allowed their (linear) source criteria to have different values at every level of recognition confidence in a mixture signal-detection model, and fits to data showed that the source criteria were more closely bunched together with higher item confidence. Finally, Klauer and Kellen (2010) found that it was critical to include a "consistency principle" in their multinomial model of recognition and source memory such that high item confidence and high source confidence tended to co-occur. Thus, the existing literature clearly supports the notion that source confidence criteria change across different levels of recognition confidence.

The BG model makes clear predictions for the difference between recognition memory zROC slopes for items receiving correct versus incorrect source attributions. The first panel of Figure 2 shows the predicted slope difference for 20,000 randomly sampled parameter sets in the bivariate Gaussian model. Parameters were chosen to represent a broad range of variation in memory performance and decision biases (see the caption for Figure 2). The *y*-axis is the slope difference, with negative values indicating a lower slope for the source-correct zROC function than the source-incorrect function. The *x*-axis is the strength of correlation between item and source evidence as an absolute value (e.g., .5 on the plot means that the correlation was .5 for Source A items and −.5 for Source B items). At low and medium values of the item-source correlation, all of the parameter sets had slope differences very close to zero. For very high correlations, the most extreme parameter sets yielded slope differences of about .15 in absolute value, but these favor the correct and incorrect source judgments equally often. Thus, the BG model generally predicts similar slopes for source-correct and source-incorrect zROC functions, and comparing the two should provide a rigorous test of the model.

Gaussian distributions produce linear zROC functions, a pattern that has been observed in numerous recognition memory experiments (e.g., Glanzer et al., 1999; Ratcliff et al., 1992). Theorists have claimed that the bivariate Gaussian model also predicts linear zROC functions in a source task (e.g., Hilford, Glanzer, Kim, & DeCarlo, 2002), but this is not true when the source criteria accommodate the fact that participants are more willing to make high-confidence source responses when item strength is high (Starns et al., 2013). As noted by Hautus et al. (2008), even a model with Gaussian distributions produces u-shaped zROC functions if the source confidence criteria converge as item strength increases. The u-shape emerges because words with low item strength rarely get high-confidence source ratings (i.e., the source criteria are widely dispersed), whereas words with high item strength commonly get high-confidence source ratings (i.e., the source criteria are compact). Coupled with the correlation between item and source evidence, this pattern means that the lower confidence source ratings—that is, the middle points of the zROC function—are predominantly made for items with low source discriminability, whereas the high-confidence source ratings—that is, the endpoints of the zROC function—are predominantly made for items with high source discriminability. Thus, the middle points are closer to chance than the end points, producing a "u" shape.

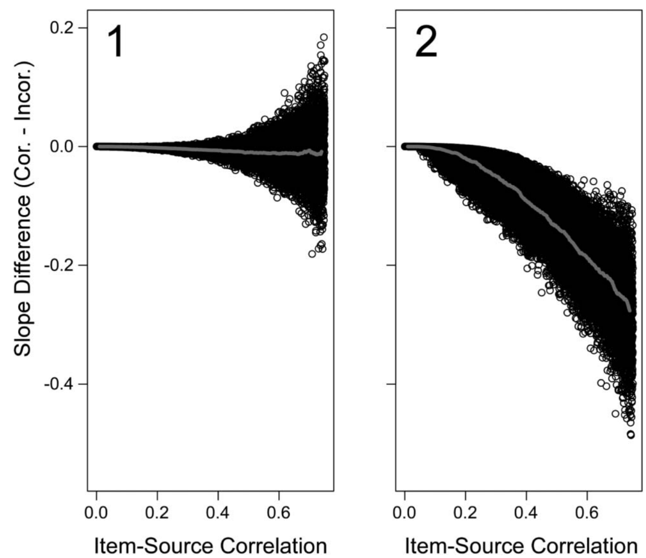

Figure 2. Predicted difference in item memory zROC slope for source-correct (Cor.) and source-incorrect (Incor.) items plotted against the correlation between item and source evidence. The gray path in each plot shows the average slope effect at each correlation value with a window of .02 (i.e., all simulation runs that used a correlation parameter between .01 below and .01 above the target correlation were included in the average). Panel 1 shows results from 20,000 simulations of the bivariate Gaussian model, and Panel 2 shows 20,000 simulations of the bounded bivariate Gaussian model. Parameter values for each simulation were drawn randomly from uniform distributions with the following ranges: from .4 to 3 for the distance between means of the new and old distributions on the item dimension, from .2 to 1.5 for the distance between the means of the Source A and Source B distributions on the source dimension, from 1 to 1.5 for the standard deviation in evidence for the studied items (independently sampled for each dimension), from 0 to .75 for the absolute value of the correlation between item and source evidence (with positive values for Source A and negative for Source B), from −.5 to .5 for the distance of the center criterion on the item dimension from the midpoint of the target and lure distributions, from .1 to .75 for the distance between each confidence criterion and the adjacent criterion, and from −1 to 1 for the position of the Source A/Source B criterion (because source responses were only classified as "correct" and "incorrect" for these simulations, the positions of the source confidence criteria did not matter).

Empirically, source zROC functions that include all of the studied items are consistently u-shaped (e.g., Glanzer, Hilford, & Kim, 2004), matching the predictions of the Gaussian model with converging source criteria. In contrast, "refined" functions that only include items recognized with high confidence are linear (e.g., Slotnick & Dodson, 2005). Results from associative recognition tasks show a similar pattern (Kelley & Wixted, 2001; Mickes, Johnson, & Wixted, 2010). Our version of the BG model predicts that refined functions should be linear, because these functions do not mix high- and low-strength items that were assessed with different source confidence criteria.

The Dual Process (DP) Model

Yonelinas (1994) developed a dual process model for ROC functions in which familiarity is a continuous strength signal that overlaps for targets and lures and recollection is a threshold process whereby the participant either retrieves or fails to retrieve qualitative details surrounding the learning event. Familiarity follows a signal detection model like the one assumed for the continuous approach described above, except that the target and lure familiarity distributions have the same variance. Thus, familiarity-based zROC functions are linear with a slope equal to 1. For recognition, adding recollection produces functions that are u-shaped with slopes less than 1, with a more pronounced u-shape and lower slope when recollection plays a larger role relative to familiarity. For source memory, recollection again produces a u-shape, but the slope should remain equal to 1 if the two sources are equal in strength (Yonelinas, 1999). Here, we extend the model to accommodate both item recognition and source confidence ratings, and we call the resulting model the bivariate dual process (BDP) model.

For the recollection process, we retain the key assumptions of Yonelinas's (1994) model: (1) recollection is a threshold process that either succeeds or fails on each trial, (2) when it succeeds, recollection produces an accurate response at the highest confidence level, and (3) the probability of recollection is independent of an item's familiarity value. Recollection can include details that specify the source of an item—such as recollecting that the item was heard in a male voice—and/or details that are not source-specific—such as recollecting that the word came after "boat" on the study list. Therefore, the model includes two recollection parameters: R_I is the overall probability that recollection will succeed for a target item, and R_S is the probability that the recollected information will include source details. For example, if $R_I = .4$ and $R_S = .5$, then source recollection will succeed for 20% of the studied items ($.4 \times .5 = .20$).

The familiarity process follows the same model shown in Figure 1, except that the target and lure distributions have the same variance in both dimensions ($\sigma_I = \sigma_S = 1$ for Source A and Source B items). The model assumes that familiarity can support both item and source discrimination. Yonelinas (1999) argued that source discrimination relies primarily on recollection when the alternative sources are equal in item strength, and this is why overall source zROCs are u-shaped.[1] More recently, however, dual process theorists have rejected the possibility that source tests uniquely rely on recollection even with equal-strength sources (Parks & Yonelinas, 2007), and a neurally inspired dual-process model also assumes that familiarity can support source discrimination (Elfman, Parks, & Yonelinas, 2008; Norman & O'Reilly, 2003). Therefore, we allowed the model to freely estimate the relative contribution of recollection and familiarity instead of assuming a priori that recollection would be dominant.

As in the Yonelinas (1994) model, in our BDP model a subset of the memory decisions are based on recollection and the remaining decisions are based on familiarity. Specifically, when recollection fails altogether ($1 - R_I$), the item and source confidence ratings are based on the test word's level of item and source familiarity, respectively. When recollection succeeds but none of the recollected details are source-specific [$R_I \times (1 - R_S)$], the item confidence rating is always a high-confidence "old" response and the source confidence rating is based on source familiarity. Recollection and familiarity are independent, as in the original Yonelinas (1994) model, so source familiarity follows the same distribution regardless of whether item recollection succeeds or fails. When recollection succeeds and source-specific information is part of the recollected details ($R_I \times R_S$), the item rating is high confidence "old" and the source decision is a correct response at the highest confidence level.

Parks and Yonelinas (2007) argued that source-specific familiarity plays a large role for refined source zROC functions. Our BDP model includes a correlation between item and source familiarity, which is consistent with the claim that selecting targets that are high in item familiarity "would lead to an apparent increase in the extent to which familiarity contributed to the source ROC" (Parks & Yonelinas, 2007, p. 195). Parks and Yonelinas introduced this mechanism to explain the finding that refined source zROC functions are more linear than functions including all of the studied items, because increasing familiarity-based source discriminability attenuates the u-shape of the function if recollection is held constant. An alternative view is that performance for refined source zROCs should be primarily based on source recollection because these functions exclude lower confidence, familiarity-based responses (Wixted, 2007). Developing an integrated model to jointly fit recognition and source data will allow us to directly address these alternatives in fits to data.

The difference in slope between source-correct and source-incorrect recognition zROC functions informs the issue of whether source decisions for high-strength items are primarily driven by recollection or familiarity. Figure 3 shows predictions from the BDP model for one parameter set in which source performance is driven primarily by recollection with little role for familiarity (Panel 1) and another parameter set in which source recollection is lower and source familiarity can be highly discriminative when item familiarity is also high (Panel 2). The top plot in each panel shows recognition memory zROC functions for targets receiving correct and incorrect source judgments, and the bottom plot shows the source memory zROC function for targets that were recognized at the highest confidence level (i.e., the refined source function). The diagonal line across each plot shows chance performance (hit rate = false-alarm rate), and functions farther from this diagonal indicate higher performance levels.

Figure 3, Panel 1, demonstrates that the BDP model predicts a lower zROC slope for the source-correct function than the source-incorrect function when recollection drives source performance. The slope difference arises because source-correct items have higher levels of recollection but similar levels of familiarity compared to source-incorrect items. The parameters for Panel 1 reflect the assumption that familiarity makes little or no contribution to source performance (Onyper et al., 2010; Wixted, 2007), so source responses based on familiarity are roughly evenly divided between correct and incorrect responses regardless of the level of item familiarity. In other words, source-correct and source-incorrect items are associated with similar levels of item familiarity. In contrast, selecting source-incorrect items eliminates all of the trials

[1] In our bivariate version of the model, converging source criteria also contribute to the u-shape, as explained previously.

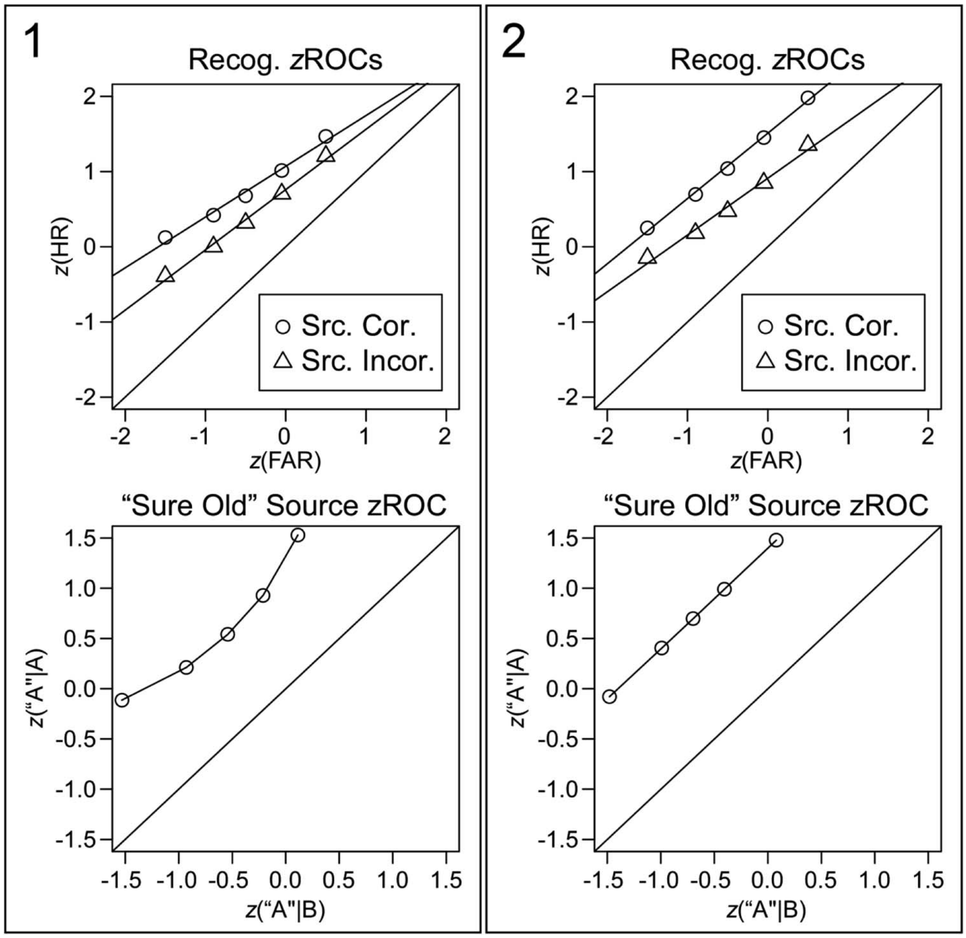

Figure 3. Predictions of the bivariate dual process model when source performance is primarily driven by recollection (Panel 1) or primarily driven by familiarity (Panel 2). The top plot in each panel shows recognition zROCs for items that were attributed to the correct source (Src. Cor.) versus the incorrect source (Src. Incor.), and the bottom plot shows source *z*-transformed receiver operating characteristic (zROC) functions for targets that were recognized with high confidence (i.e., the refined source zROC). For Panel 1, the means of the familiarity distributions for Source A and Source B items were both 0.6 on the item dimension and 0.05 and −0.05 on the source dimension. The correlation between item and source evidence was .05 and −.05 for Sources A and B. The probability of recollecting the learning event (R_I) was .35, and source-specific details were retrieved for half of these recollected items (R_S = .5). Panel 2 used the same parameter values with the following exceptions: The means of the familiarity distributions were 0.5 and −0.5 for Source A and B items, the correlation between source and item evidence were .5 and −.5, and the proportion of recollected items that included source-specific details (R_S) was .05. z(FAR) = *z*-score of the false-alarm rate; z(HR) = *z*-score of the hit rate; z("A"|B) = *z*-score of the proportion of Source B items attributed to Source A; z("A"|A) = *z*-score of the proportion of Source A items attributed to Source A.

in which source recollection succeeded, whereas selecting source-correct items retains all of these trials. Thus, recollection is higher overall for source-correct trials, and the difference is substantial if source-specific recollection is common. Finally, the refined source zROC function is u-shaped in Panel 1 because recollection is the primary process driving source discrimination for high-strength items.

Figure 3, Panel 2, shows that the BDP model predicts the opposite pattern for recognition zROC slopes under the assumption that source recollection is low and source familiarity is highly discriminative for items with high item familiarity. In this case, source-correct and source-incorrect items have similar levels of recollection, but source-correct items are higher in item familiarity. That is, high-familiarity items are often attributed to the correct source, whereas low-familiarity items are associated with much poorer source performance. As a result, selecting source-correct items also selects items that tend to be high in item familiarity. However, source recollection is generally low, so retaining the trials with successful source recollection for the source-correct items does not result is much added recollection compared to eliminating these trials for the source-incorrect items. With familiarity driving source identification for high-strength items, the refined source zROC is linear (as suggested by Parks & Yonelinas, 2007).

Figure 4 demonstrates that the pattern shown in Figure 3 does not depend on the specific parameter values used to generate predictions. The plot shows 20,000 sets of model predictions with the parameters for each run drawn from uniform distributions spanning a wide range of values (see the caption for Figure 4). For each parameter set, the predicted difference in slope between the source-correct and source-incorrect recognition zROC function is plotted against the quadratic coefficient for the refined source zROC function. A quadratic coefficient of zero indicates a linear function; positive values indicate a u-shaped function; and negative values indicate an inverted-u-shaped function. Again, the model can predict either positive or negative slope effects depending on the selected parameter values. With no slope effect or a positive slope effect, the model usually produces u-shaped refined zROCs, but linear functions and functions with a slight inverted-u shape are also possible.[2] However, with even a small negative slope effect, the model must produce a u-shape for the refined zROC function.

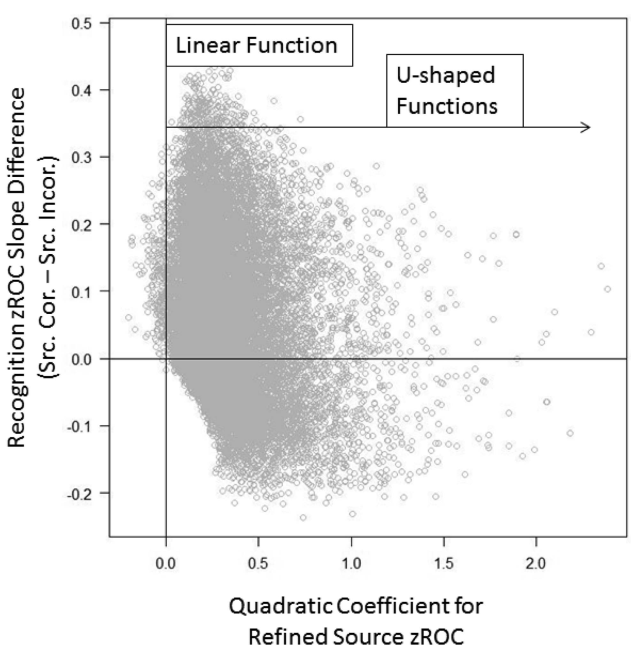

Figure 4. Predictions form 20,000 runs of the bivariate dual process model across a wide range of parameter values. The *y*-axis is the difference in slope between the source-correct (Src. Cor.) and source-incorrect (Src. Incor.) recognition *z*-transformed receiver operating characteristic (zROC) functions, and the *x*-axis is the quadratic coefficient for the refined source zROC. Each point shows the model predictions from one randomly selected set of parameter values. Parameters were randomly selected from uniform distributions with the following ranges: from .2 to 1.5 for the distance between means of the new and old distributions on the item familiarity dimension, from .1 to 2 for the distance between the means of the Source A and Source B distributions on the source familiarity dimension, from 0 to .75 for the absolute value of the correlation between item and source evidence (with positive values for Source A and negative for Source B), from 0 to 2 for the position of the center (“old”/“new”) recognition criterion, from −.5 to .5 for the position of the center source criterion, from .3 to 1 for the distance between each confidence criterion and the adjacent criterion, from .1 to .7 for the proportion of items for which item recollection succeeds, and from .1 to .7 for the proportion of times recollection includes source-specific information.

Figures 3 and 4 suggest that comparing source-correct and source-incorrect zROC slopes will be a good test of the BDP model. Although the model can accommodate slope effects in either direction, the parameter adjustments required to make this change affect predictions for other aspects of the data. If the source-correct function has a lower slope than the source-incorrect function, then the model predicts that the refined source zROC must be u-shaped. If the source-correct function has a higher slope, then linear functions are possible for the refined source zROC. Finding a linear refined source zROC along with a lower slope for the source correct function would clearly violate the model’s predictions.

The Bounded Bivariate Gaussian (BBG) Model

The model shown in Figure 1 has an unfortunate property: because the source distributions are slanted in opposite directions, they actually cross at some point as item strength decreases. For example, the model could predict that if one selects out targets that received a high confidence “new” response, then the Source A items should be consistently attributed to Source B, and vice versa (Hautus et al., 2008). Of course, this pattern is not observed in empirical data. Hautus et al. (2008) worked around this property by introducing a source-guessing process for unrecognized items (also see Slotnick & Dodson, 2005). Here, we address the issue by directly eliminating the crossover in the distributions. Specifically, we replaced the standard bivariate distributions with what we call “bounded” bivariate distributions, producing the *bounded bivariate Gaussian* (BBG) model (see Figure 5).

To understand the difference between the BG and BBG models, consider the mean of the source evidence distribution conditionalized on a particular level of item evidence. At the item-evidence mean, the conditional source-evidence mean is positive for Source A items and Negative for Source B. In the BG model, the conditionalized mean decreases as item evidence decreases for Source A items (i.e., a positive correlation) and increases as item evidence decreases for Source B items (i.e., a negative correlation). At one point on the item continuum, the conditionalized source mean will equal zero, and we call this the *crossover point*. For item evidence values above this crossover point, the BBG model is identical to the BG model. For all item evidence values below this crossover point, the BBG model assumes that the conditional source mean is zero; that is, source discrimination is at chance. This implements the psychological intuition that very weak learning should be insufficient to establish source information, and that no level of

[2] Readers might find it strange that inverted-u functions are possible, given that the Gaussian familiarity distributions should produce a linear function and any role for recollection introduces a u-shape. The explanation is that the familiarity distributions for the refined source zROC are not necessarily Gaussian—they are the portion of a bivariate Gaussian distribution that falls above the highest recognition criterion (because the refined zROC includes only “sure old” responses). With zero correlation between the two dimensions, the selected source distribution is Gaussian, but large correlations can produce skewed distributions that impart an inverted-u shape to the zROC function. The *overall* source zROCs predicted by the model can never have an inverted-u shape, but conditional zROCs can.

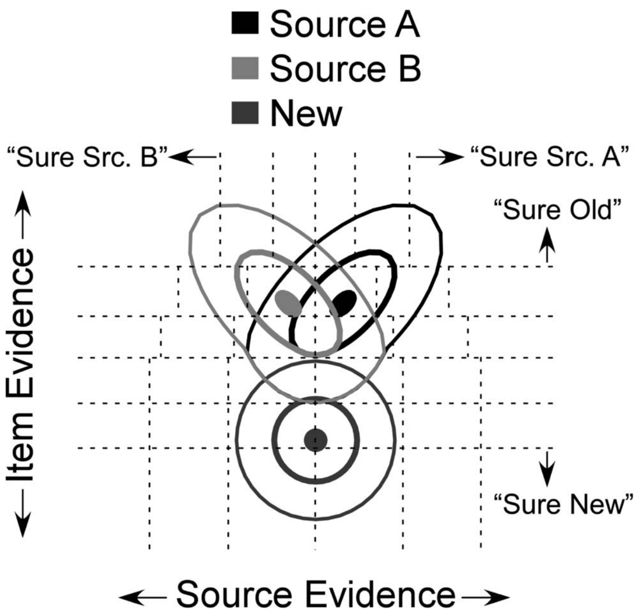

Figure 5. The bounded bivariate Gaussian model of recognition and source confidence ratings. The sets of ovals represent bounded bivariate Gaussian distributions for items studied in Source A (Src. A), items Studied in Source B (Src. B), and items appearing for the first time at test (New). The solid dots show the means of each distribution, and the surrounding lines are equal-density contours with thicker lines for higher probability densities. Five criteria on the item dimension define regions associated with six item-confidence ratings. On the source dimension, the model has separate source criteria for "New" responses and for each of the three confidence levels of "Old" responses.

learning should systematically establish false source information. The Appendix defines the probability-density function for a bounded bivariate Gaussian distribution.

Interestingly, eliminating the psychologically implausible crossover in the source evidence distributions dramatically changes the model predictions for the difference between source-correct and source-incorrect zROC slopes. The second panel of Figure 2 shows slope differences from 20,000 randomly sampled parameter sets in the BBG model (using the same parameter ranges as the BG model in Panel 1). The BBG model predicts no slope difference if item and source evidence are uncorrelated. In contrast to the BG model, as the correlation increases, the BBG model consistently predicts that the source-correct function will have a lower slope than the source-incorrect function (leading to negative differences). The size of the slope difference can be quite large for moderate to high correlations.

Figure 2 shows that the BBG model is tightly constrained to predict that the item recognition zROC has a lower slope for source-correct versus source-incorrect items whenever recognition and source performance are related (and this relationship is always observed empirically; e.g., Glanzer et al., 2004). The logic behind this prediction is intuitive: Targets that have very low item strength are associated with chance performance on the source question, so they will be evenly divided into the source-correct and source-incorrect categories. Targets that are very high in item strength are consistently attributed to the correct source, so the source-correct category of items will span the full continuum of item strength. In contrast, the source-incorrect category will largely exclude the highest-strength items. Thus, the source-correct function has a lower slope because item strength is more variable for these items.

Like the BG model, the BBG model predicts a u-shaped overall source zROC function and a linear refined function. The u-shape is partially driven by the converging source criteria, as explained for the BG model. However, the evidence distributions can contribute to the u-shape for the BBG model as well. In the BBG model, the source distributions are Gaussian at any given level of item evidence, but the marginal source distribution across all items can be non-Gaussian. If a high proportion of items fall below the cutoff point on the item dimension where the source distributions merge, then the marginal source distribution will be "bunched up" at low performance levels. That is, there will be a higher proportion of items around a source strength of zero than one would expect for a Gaussian distribution. More than convenience motivates these skewed distributions for source judgments: Shimamura and Wickens (2009) arrived at similar distributional assumptions based on consideration of the properties of the medial temporal lobe. A natural consequence of these skewed distributions is that they impart a u-shape to the zROC function, similar to the effect of adding random guesses on a proportion of trials (Hautus et al., 2008; Slotnick & Dodson, 2005).[3]

We also fit a version of the BDP model using bounded bivariate distributions for the familiarity process, and we call this the *bounded bivariate dual process* (BBDP) model. This version of the dual process model can still predict slope differences in either direction based on whether source discrimination is driven by recollection or familiarity, as displayed in Figure 3 for the BDP model. As detailed below, the BDP and BBDP models had very similar fits and parameter values.

Summary

We have described models that make distinct predictions about the slope of the item recognition zROC function for items receiving correct versus incorrect source judgments. The bivariate Gaussian (BG) model predicts little or no difference in slope (see Figure 2, Panel 1). A modification to this model that eliminates below-chance source discrimination for all levels of item strength—the bounded bivariate Gaussian (BBG) model—predicts a lower slope for source-correct than source-incorrect items (see Figure 2, Panel 2). The bivariate and bounded bivariate versions of the dual process model (BDP and BBDP models, respectively) can predict slope differences in either direction depending on assumptions about memory processes. If source discrimination relies primarily on the high-threshold recollection process, then the recognition zROCs should have a shallower slope for source-correct than source-incorrect items and the refined source zROC must be u-shaped (see Figure 3, Panel 1). If familiarity drives source

[3] The BBG model is similar, but not identical, to the proposal that participants make random source guesses for any item they fail to recognize (Hautus et al., 2008; Slotnick & Dodson, 2005). One critical difference is that the point where the bounded distributions converge is not determined by a participant's criterion for making an "old" response, so in the BBG model, it is possible for "new" responses to have above chance source discrimination if the recognition criterion is sufficiently conservative (as observed by Starns, Hicks, Brown, & Martin, 2008).

performance for high-strength items, then the recognition zROCs should have a steeper slope for source-correct than source-incorrect items and the source zROC can be linear (see Figure 3, Panel 2). In what follows, we compare these predictions to empirical data.

Method

Data Sets

We re-analyzed data from Yonelinas (1999, Experiment 2) as well as from Starns et al. (2013, Experiments 2a and 2b), although the latter data were not used for model fitting because the recognition and source judgments were made on separate tests. We also analyzed data from three experiments that are fully reported here for the first time. The methods for these experiments are described below. Data for the new items from these experiments contributed to Figure 7 in Starns et al. (2013), in which they appeared as Data Sets 6, 9, and 11. Data for the studied items are reported here for the first time, and we also report the first modeling attempts for these data.

Participants

Experiment 1 included 23 undergraduate volunteers from the University of Massachusetts Amherst who completed the experiment for credit in their psychology courses. Experiments 2 and 3 included 27 and 28 participants from the same pool. None of the participants took part in more than one experiment. Subjects who were near chance performance for item recognition were excluded from the data analysis, leaving 16, 23, and 22 subjects in the final pool for each experiment.

Design

For all three experiments, the only independent variable was item type with the levels male, female, and new. Item type was manipulated within subjects. All three experiments had equal learning strength for male and female items and used a "neutral" test that did not explicitly bias participants to attribute items to one of the sources.

Materials and Procedure

Experiments 1–3 all had a study list composed of 160 common nouns with half randomly assigned to be presented in a male voice and the rest presented in a female voice. Words were presented in a unique random order for each subject. Each study word appeared on the computer screen for 3 s and was accompanied by an auditory recording of the word. There were also five primacy and five recency stimuli that were not tested.

The test procedures were also very similar across experiments. In Experiment 1, all 160 of the studied words appeared at test mixed with 80 new words. Test words appeared one at a time on the center of the computer screen, and no audio files were played during the test phase. For each memory probe, subjects were asked to rate their confidence that the word had been studied (6 = "sure old," 5 = "probably old," 4 = "maybe old," 3 = "maybe new," 2 = "probably new," 1 = "sure new") and then, regardless of their old-new confidence, to rate their confidence that the word had been presented in a male voice (6 = "sure male," 5 = "probably male," 4 = "maybe male," 3 = "maybe female," 2 = "probably female," 1 = "sure female"). The ratings were made by pressing the number keys on the computer keyboard, and each test word remained on the screen until both ratings were made. Experiment 2 was identical, except that participants were informed that they would earn 2 points for each correct response and lose 3 points for each error. Each point translated to one ticket in a drawing for a $50 prize that was held after all the sessions were completed. Experiment 3 was identical to Experiments 1 and 2 during the study phase, but at test half of the male and female items from the study phase were randomly selected to appear (i.e., 40 of each), along with 40 new items. Across all experiments, the study words and new items for the test were randomly selected from the same pool for each participant.

Results

zROC Slope Effect

Our primary focus was the difference between the recognition memory zROC slopes for items with correct and incorrect source attributions. These data are reported in Table 1. The first four rows show the data sets that contributed to the model fits below. For these experiments, participants made recognition and source ratings in immediate succession for each test item. All of the source-correct functions had a considerably lower slope than the source-incorrect functions. The last four rows show data from Experiments 2a and 2b in Starns et al. (2013). These experiments had separate test lists for recognition and source judgments. Again, the source-correct slope was lower for all four experiments. Thus, the slope difference was quite robust. Under the null hypothesis that the slopes are equal, the probability of observing eight negative slope differences in eight experiments is .004.

These data are most consistent with the predictions of the BBG model, which is highly constrained to predict a lower source-correct slope as long as item and source performance are correlated. The data seem to pose a substantial challenge to the BG model, which can only produce slope differences in limited circumstances (and even then the effects tend to be small). The dual

Table 1
Slope of the Recognition zROC Function for Items With Correct Versus Incorrect Source Decisions

Data set	Source correct	Source incorrect	Difference
Yonelinas (1999, Experiment 2)	.632	.853	−.221
Experiment 1	.557	.779	−.222
Experiment 2	.593	.785	−.192
Experiment 3	.525	.785	−.260
Starns et al. (2013)—Separate source and recognition tests			
Experiment 2A Weak	.625	.777	−.152
Experiment 2A Strong	.651	.807	−.156
Experiment 2B Weak	.644	.737	−.093
Experiment 2B Strong	.632	.795	−.163

Note. zROC = *z*-transformed receiver operating characteristic.

process models can produce the correct slope difference when recollection plays a large role in source responding, which suggests that the refined source zROCs will be u-shaped (see Figure 3, Panel 1, and Figure 4). Next, we compare the models in terms of direct fits to data, with a focus on whether the models can match the qualitative patterns that are consistently observed across experiments.

Model Fitting

The Appendix fully describes the parameters in each model and reports the prediction equations. The models were fit to the response frequencies in the 36 categories defined by factorially crossing the six recognition and six source confidence levels. Each item type contributed 35 degrees of freedom (*df*) to the fits (one of the response frequencies is constrained to sum to the total number of observations). With three item types (male, female, and new), the data had a total of 105 *df*. Each model had 30 free parameters. Twenty-five of the parameters were response criteria, with five criteria for item confidence ratings and 20 total source criteria—five each for items called "new," "maybe old," "probably old," and "sure old." The remaining parameters defined the memory evidence distributions: the item evidence mean (μ_I) and standard deviation (σ_I), the source evidence mean (μ_S) and standard deviation (σ_S), and the correlation between item and source evidence (ρ). The standard deviation parameters were fixed at 1 for the dual process models, but this model had additional free parameters for the overall probability that recollection would succeed for an item (R_I) and the probability that the recollected information would include source information (R_S). Given that all of our data sets had equal-strength sources, all of the memory evidence parameters had the same value for male and female items, although with an opposite sign for the μ_S and ρ parameters. In assessing the model fits, we focused on the data that clearly discriminate the competing models, namely the recognition zROC functions for source-correct and source-incorrect items and the overall and refined source zROC functions. The BDP and BBDP models provided very similar fits with a slight advantage for the BBDP model for every data set. Parameter values and visual fits were nearly identical for the BBDP and BDP models, so we report results only for the BBDP model. Table 2 shows the best fitting memory parameters from each model.

Recognition memory zROC functions. Figure 6 shows the fit of the BG model to the source-correct and source-incorrect zROC functions. For all four data sets, recognition performance was higher for source-correct than source-incorrect items, evincing a correlation between item and source evidence. Most of the functions were linear, as is usually observed for recognition zROCs (e.g., Glanzer et al., 1999; Ratcliff, McKoon, & Tindall, 1994). The functions for Experiment 1 had an inverted-u shape, which cannot be produced by any of the models under consideration. Some reaction time (RT) models can produce these functions (Ratcliff & Starns, 2009, 2013), an issue we return to in the General Discussion. As reported in Table 1, the source-correct functions have a lower slope than the source-incorrect functions, as shown by the black lines. In contrast to the data, the BG model produced essentially parallel slopes for the source-correct and source-incorrect zROCs, as shown by the gray lines. Figure 7 shows the fit of the BBG model to the recognition memory zROC functions. This model produced a very close match to the data, including the shallower slope for the source-correct than the source-incorrect functions.

Table 2
Memory Evidence Parameters and G^2 Values

Data set and model	Parameter							
	G^2 (75)	μ_I	σ_I	μ_S	σ_S	ρ	R_I	R_S
Yonelinas (1999, Experiment 2)								
BG	200	1.30	1.41	0.34	1.09	.33		
BBG	127	1.29	1.40	0.17	1.25	.56		
BBDP	123	0.61	1[a]	0.03	1[a]	.22	.33	.52
Experiment 1								
BG	338	1.39	1.57	0.42	1.15	.29		
BBG	298	1.35	1.54	0.31	1.26	.48		
BBDP	373	0.56	1[a]	0.08	1[a]	.25	.35	.46
Experiment 2								
BG	211	1.39	1.55	0.42	1.11	.32		
BBG	183	1.38	1.54	0.33	1.22	.50		
BBDP	207	0.53	1[a]	0.08	1[a]	.19	.40	.52
Experiment 3								
BG	160	1.76	1.69	0.58	1.15	.41		
BBG	118	1.72	1.66	0.47	1.39	.66		
BBDP	148	0.63	1[a]	0.02	1[a]	.29	.45	.65

Note. The G^2 values are associated with 75 degrees of freedom (*df*): 105 *df* in the data minus 30 free parameters for each model. The mean and standard deviation in item strength (μ_I and σ_I) were equal for male and female items. The mean source strength (μ_S) and item-source correlation (ρ) are reported for male items, and the corresponding values for female items were negative with the same absolute value. Male and female items had the same standard deviation in source strength (σ_S). R_I = probability of recollecting any details of studying the item; R_S = probability that source details will be part of the recollected information; BG = bivariate Gaussian; BBG = bounded bivariate Gaussian; BBDP = bounded bivariate dual process. The bivariate dual process model always fit slightly worse than the BBDP model (G^2 of 134, 380, 210, and 153 for the four data sets, respectively).
[a] This parameter was fixed.

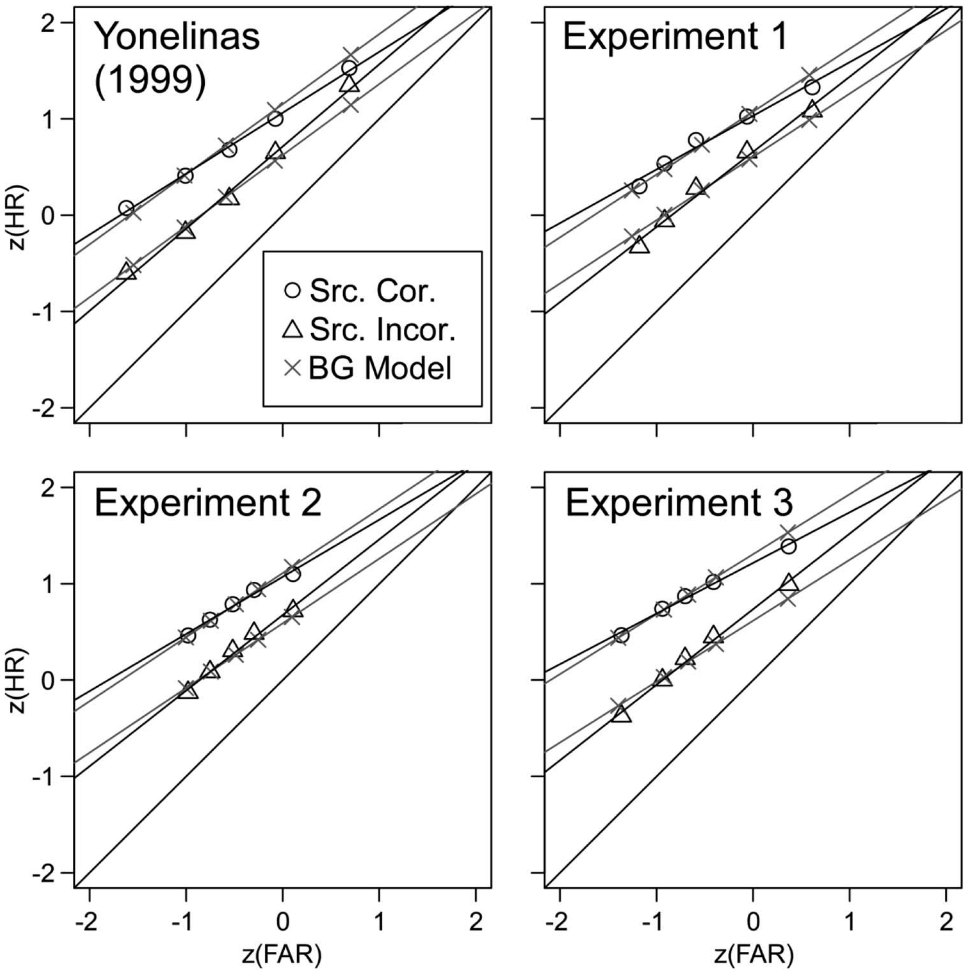

Figure 6. Fit of the bivariate Gaussian (BG) model to recognition memory *z*-transformed receiver operating characteristic (zROC) functions for source-correct (Src. Cor.) and source-incorrect (Src. Incor.) items. The black lines show the zROC slopes for the data, and the gray lines show the zROC slopes for the model predictions. Panels 1–4 show results from Yonelinas (1999) and Experiments 1–3, respectively. z(FAR) = *z*-score of the false-alarm rate; z(HR) = *z*-score of the hit rate.

The fits for the BBDP model are shown in Figure 8, which reveal that this model was also able to match the slope effect. The predicted slopes for the BBDP model were sometimes higher than the empirical slopes, especially for the source-correct functions. However, the model generally matched the qualitative patterns in the data, even if the quantitative details were sometimes not well-fit (e.g., Figure 8, Panel 2, Experiment 1). As detailed in the Introduction, recollection must play a fairly large role in the source task for the dual process model to predict lower slopes for the source-correct items. Table 2 reports the best fitting memory parameters and G^2 values for each model, and the recollection parameters do show a prominent role for recollection. The R_I values suggest that recollection succeeded for between .33 and .45 of the studied items across the data sets, and the R_S values indicate that the recollected details included source information roughly half of the time (.46–.65). In contrast, the μ_S parameters indicate that familiarity-based source discrimination was low; that is, the means of the source familiarity distributions were close to zero.

Source memory zROC functions. Figure 9 shows the fit of the BG model to the overall and refined source zROC functions for each of the four data sets. As in the previous literature (Glanzer et al., 2004; Slotnick & Dodson, 2005; Yonelinas, 1999), the overall functions are u-shaped, although the deviation from linearity is more subtle for Experiments 1 and 2. The refined functions are basically linear, and the points are more compact than the overall functions. The predicted overall functions for the BG model were also u-shaped, although the u-shape in the predicted function is often more subtle than in the data. As described in the Introduction, the u-shape is produced because the source confidence criteria become more compact as recognition confidence increases. For example, in Yonelinas's (1999) data set, the distance from the lowest criterion (separating the "Sure Female" and "Probably Female" responses) to the highest criterion (separating the "Probably Male" and "Sure Male" responses) was 4.47, 4.27, 3.36, and 1.23 for "new," "maybe old," "probably old," and "sure old"

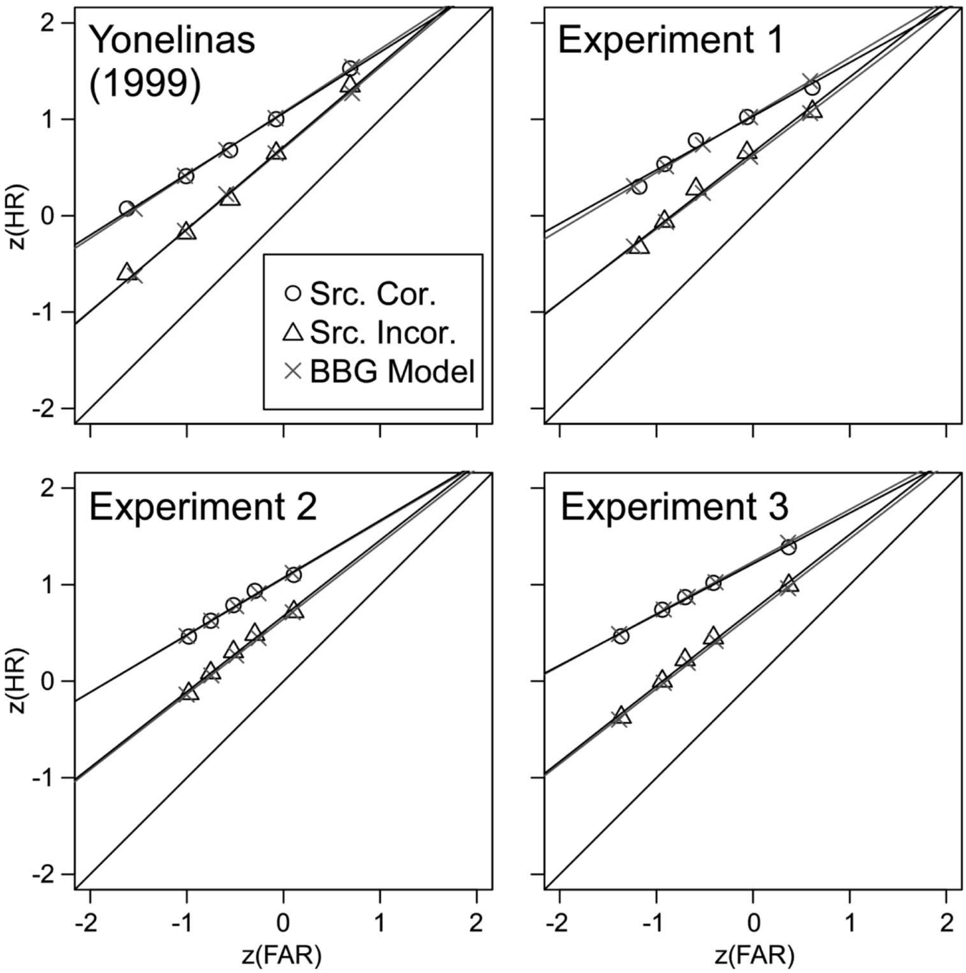

Figure 7. Fit of the bounded bivariate Gaussian (BBG) model to recognition memory *z*-transformed receiver operating characteristic (zROC) functions for source-correct (Src. Cor.) and source-incorrect (Src. Incor.) items. The black lines show the zROC slopes for the data, and the gray lines show the zROC slopes for the model predictions. Panels 1–4 show results from Yonelinas (1999) and Experiments 1–3, respectively. z(FAR) = *z*-score of the false-alarm rate; z(HR) = *z*-score of the hit rate.

responses, respectively. This converging criteria pattern was observed for all of the data sets and for all of the models under investigation. For the refined functions, the model matched the observed linear functions because source decisions for these items were made with a single set of source confidence criteria (the criteria for the highest recognition confidence level).

Although the BG model generally matched the shapes of the overall and refined zROC functions, the model consistently under-predicted performance for the refined source zROCs. These mispredictions suggest that the estimated correlation between item and source evidence was too weak in the model, so selecting out words with high item strength did not improve source memory as much in the model as in the data. Increasing the correlation would improve performance for the refined source zROC, but would also accentuate the problem of below-chance source discrimination for low-strength items that was noted in the Introduction. Therefore, the psychologically implausible crossover in the source distributions appears to prevent the BG model from accommodating the data.

Figure 10 shows the source zROC fits for the BBG model. This model successfully matched all aspects of the data, including the u-shape for the overall zROC functions and the linearity of the refined zROC functions. Moreover, the BBG model closely matched the refined source performance, in contrast to the consistent under-prediction seen with the BG model. Table 2 reveals that the item-source correlations (ρ) were higher for the BBG model than the BG model for every data set. As a result of using bounded distributions, the BBG model can increase the relationship between item and source performance without exacerbating the psychologically implausible crossover in the source distributions. This property brings the continuous approach into much better accord with the data.

Figure 11 shows the source zROC functions for the BBDP model. Like the other models, the BBDP model matched the u-shape in the overall zROC functions. However, this model produced u-shaped refined zROC functions in contrast to the linear empirical functions, as evinced by consistent underestimates for

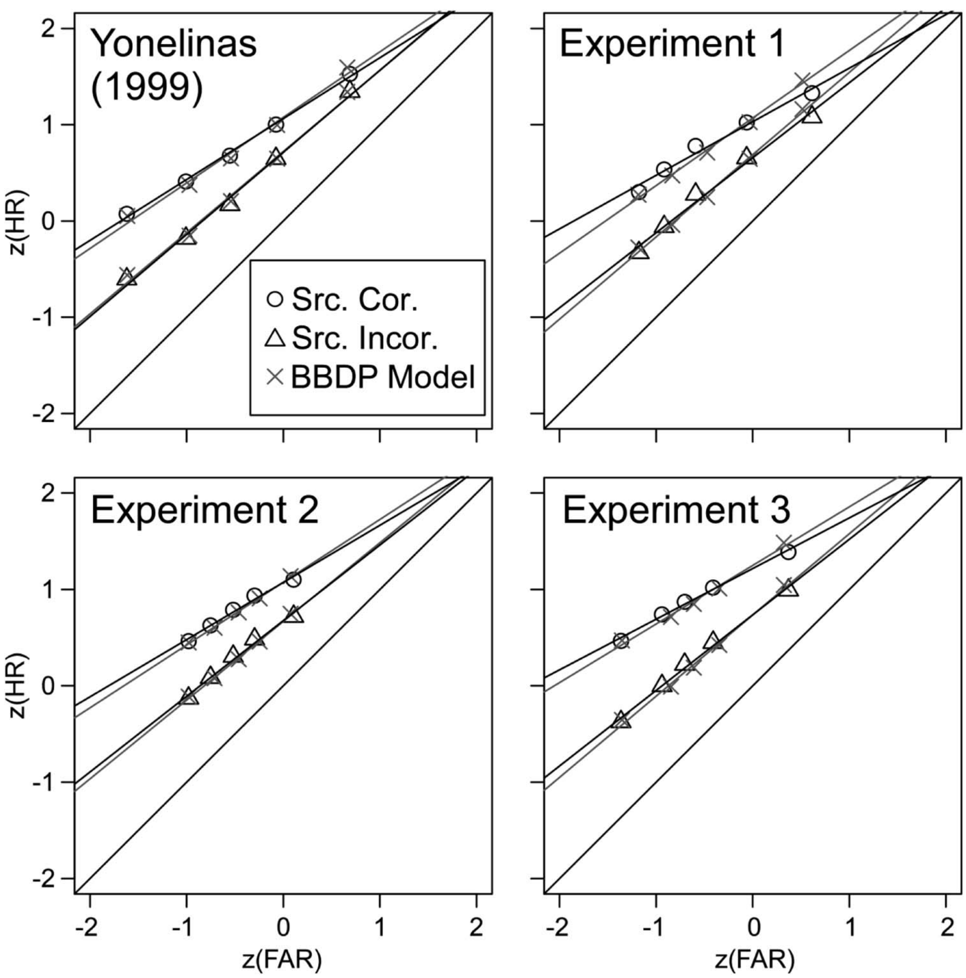

Figure 8. Fit of the bounded bivariate dual process (BBDP) model to recognition memory *z*-transformed receiver operating characteristic (zROC) functions for source-correct (Src. Cor.) and source-incorrect (Src. Incor.) items. The black lines show the zROC slopes for the data, and the gray lines show the zROC slopes for the model predictions. Panels 1–4 show results from Yonelinas (1999) and Experiments 1–3, respectively. z(FAR) = *z*-score of the false-alarm rate; z(HR) = *z*-score of the hit rate.

the middle zROC points. These mis-predictions are most apparent for Yonelinas's (1999) data set and Experiments 1 and 3, with a fairly good match to the data in Experiment 2. Again, the model predicts a u-shaped refined function as long as recollection plays a large role in source discrimination, and the model cannot match the source-correct and source-incorrect zROC slopes unless recollection is the primary basis of source discrimination.

Goodness of Fit

The recognition and source zROC fits suggest that the BBG model outperformed the other two models. The G^2 fit statistics reported in Table 2 are consistent with this conclusion. The BBG model had a lower G^2 value than the other models for three of the four data sets, and for the remaining data set the BBG and BBDP models produced a very similar fit with a worse fit for the BG model. All of the models have the same number of parameters, so fit metrics that adjust for model complexity in terms of the number of free parameters would produce the same conclusions as the G^2 values (Akaike, 1973; Schwarz, 1978). All of the G^2 values are above the critical value for significance at the .05 level (96), but they are consistent with the G^2 values reported in past fits to joint recognition and source confidence ratings (e.g., Hautus et al., 2008; Klauer & Kellen, 2010).

Fit metrics that assess complexity in terms of functional form might not produce the same conclusions as G^2 (e.g., Pitt, Myung, & Zhang, 2002); however, we note that the predictions of the BBG model were highly constrained for these data. Most critically, the BBG model cannot match data in which the recognition zROC slope is higher for source-correct than source-incorrect items, and it also cannot match a null effect for any data set with a moderate to strong correlation between item and source performance (see Figure 2, Panel 2). Thus, the BBG model was successful not because of its flexibility, but because the data consistently conformed to the highly constrained predictions of this model.

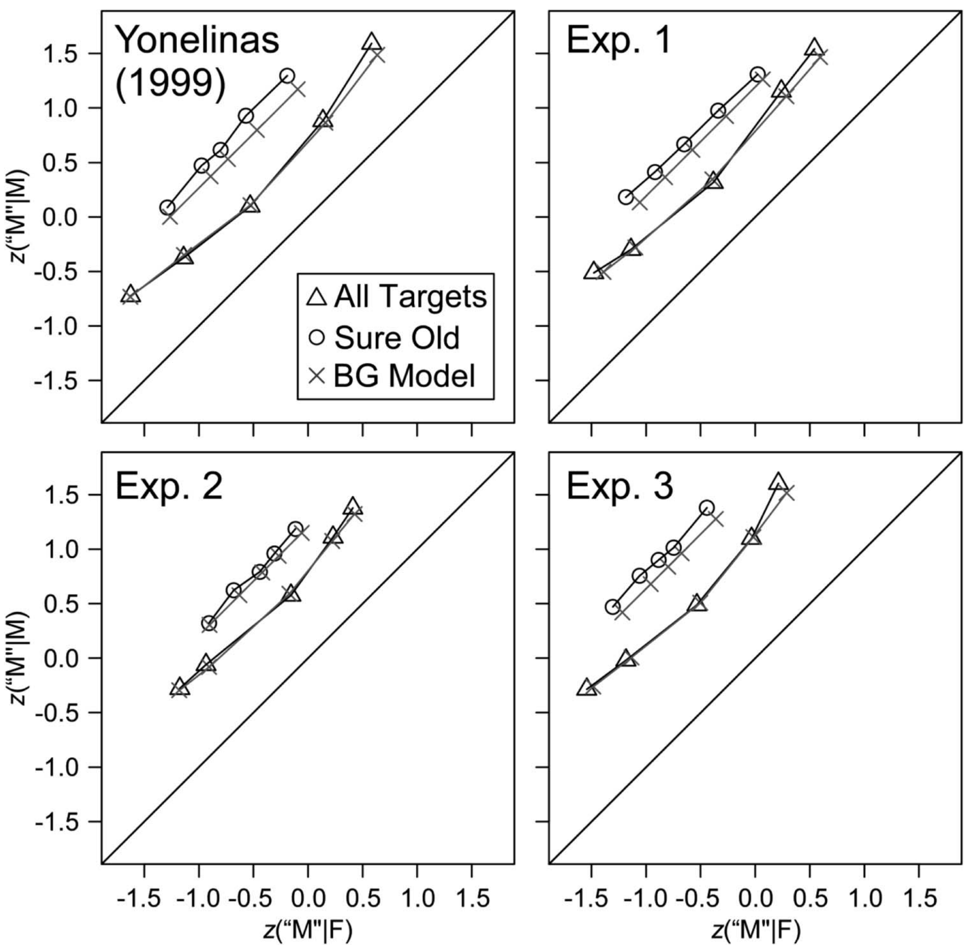

Figure 9. Fit of the bivariate Gaussian (BG) model to the source *z*-transformed receiver operating characteristic (zROC) functions for all target items and just for targets receiving a "Sure Old" response (i.e., the refined source function). Panels 1–4 show results from Yonelinas (1999) and Experiments 1–3, respectively. z("M"|F) = *z*-score of the proportion of female items attributed to the male voice; z("M"|M) = *z*-score of the proportion of male items attributed to the male voice.

General Discussion

Three qualitative patterns were observed in every data set we analyzed: (1) the recognition memory zROC slope was lower for items receiving correct than incorrect source responses, (2) the overall source zROC was u-shaped, and (3) the refined source zROC was linear. We evaluated unequal variance and dual process models in terms of their ability to match these patterns. Only the bounded bivariate Gaussian (BBG) model could produce all of these patterns simultaneously. The bivariate Gaussian (BG) model could not produce the difference between source-correct and source-incorrect zROC slopes, and it consistently under-predicted performance for the refined source zROC functions. The bounded bivariate dual-process (BBDP) model incorrectly predicted that the refined source zROC functions should be u-shaped, because this model had to propose that source performance depended heavily on recollection to match the difference in source-correct and source-incorrect zROC slopes.[4]

Wixted (2007) argued that the dual process account is unable to explain why refined source zROC functions are linear, but Parks and Yonelinas (2007) attempted to account for this finding by proposing that source familiarity can be highly discriminable when item familiarity is high (pp. 194–195). We developed an explicit model capable of implementing the Parks and Yonelinas account by introducing a correlation between item and source familiarity. This model shows that linear refined zROCs are possible (see Figure 3, Panel 2), but only when the source-correct zROC slope is at least as steep as the source-incorrect slope (see Figure 4). That is, assuming that correct source responses are often based on source familiarity when item familiarity is high also means that items with correct source responses tend to be more familiar than items with incorrect source responses, and higher familiarity for the source-correct items produces a zROC slope closer to 1 in the model. We reported analyses of eight data sets all showing that

[4] As mentioned previously, fits of the dual process model with unbounded familiarity distributions were always similar to but slightly worse than the BBDP model.

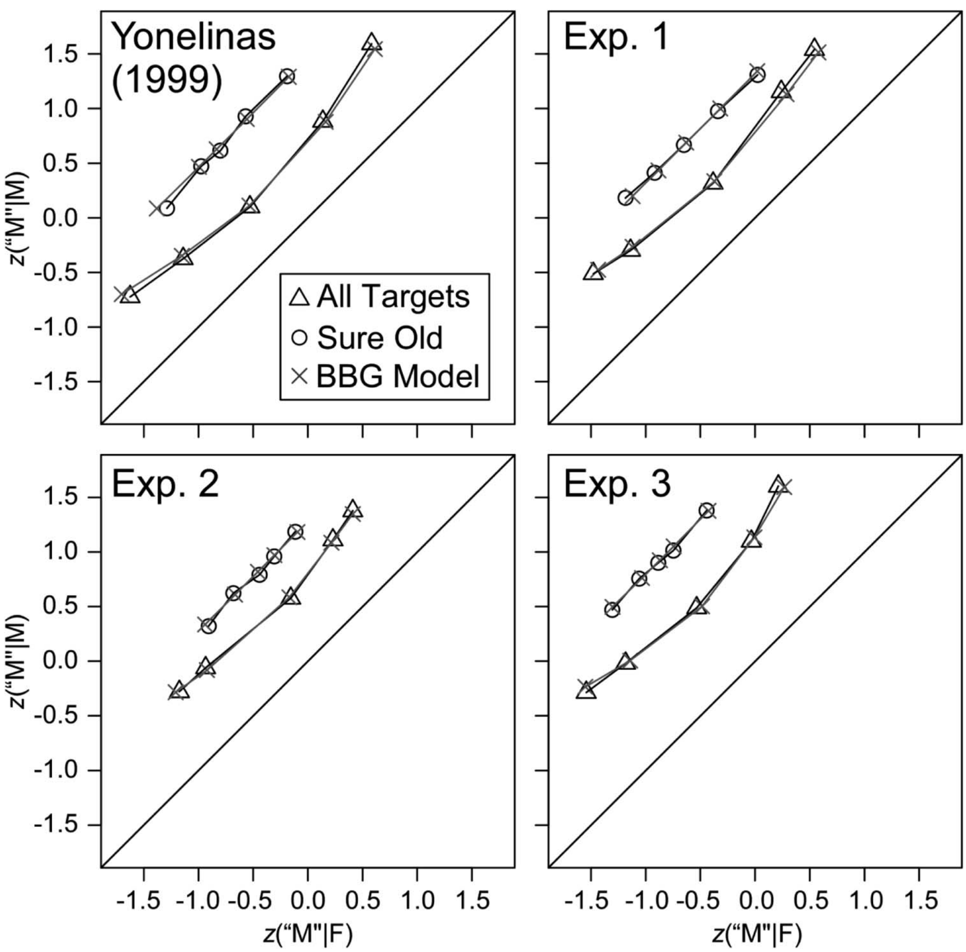

Figure 10. Fit of the bounded bivariate Gaussian (BBG) model to the source *z*-transformed receiver operating characteristic (zROC) functions for all target items and just for targets receiving a "Sure Old" response (i.e., the refined source function). Panels 1–4 show results from Yonelinas (1999) and Experiments 1–3, respectively. z("M"|F) = *z*-score of the proportion of female items attributed to the male voice; z("M"|M) = *z*-score of the proportion of male items attributed to the male voice.

the source-correct slope was substantially lower than the source-incorrect slope (see Table 1). Thus, Parks and Yonelinas's account is not viable, and the dual process model has no apparent explanation for the fact that refined source zROC functions are linear. Linear refined functions were observed for every data set herein and in past reports (e.g., Slotnick & Dodson, 2005).

Decision Biases Affect zROC Shape

One important conclusion supported by the current data is that the slope and shape of zROC functions can be influenced by decision processes as well as memory processes. Decision effects can distort conclusions about memory in any model that links zROC characteristics to memory processes, including the dual process, unequal variance, and mixture signal detection models. For example, Parks and Yonelinas (2007) claimed that the u-shape that is consistently observed in overall source zROC functions "directly conflicts with the Gaussian assumption underlying the UVSD [unequal variance signal detection] model" (p. 192). Although this is true with standard assumptions about decision criteria, our results show that decision mechanisms can produce u-shaped functions even when the underlying evidence distributions are Gaussian. In the bivariate Gaussian model, the marginal source distributions are univariate Gaussian distributions identical to those assumed in the UVSD model. Nevertheless, the BG model predicted u-shaped overall source zROCs. The u-shape was produced by changes in the source confidence criteria across different levels of item strength. Importantly, other aspects of the data show that the source criteria change in this way, such as the fact that lures receive more high-confidence source attributions when they are falsely recognized with higher confidence levels (Starns et al., 2013).

Aggregation Issues

The critical failure of the dual process model was that the model consistently produced u-shaped refined source functions whereas

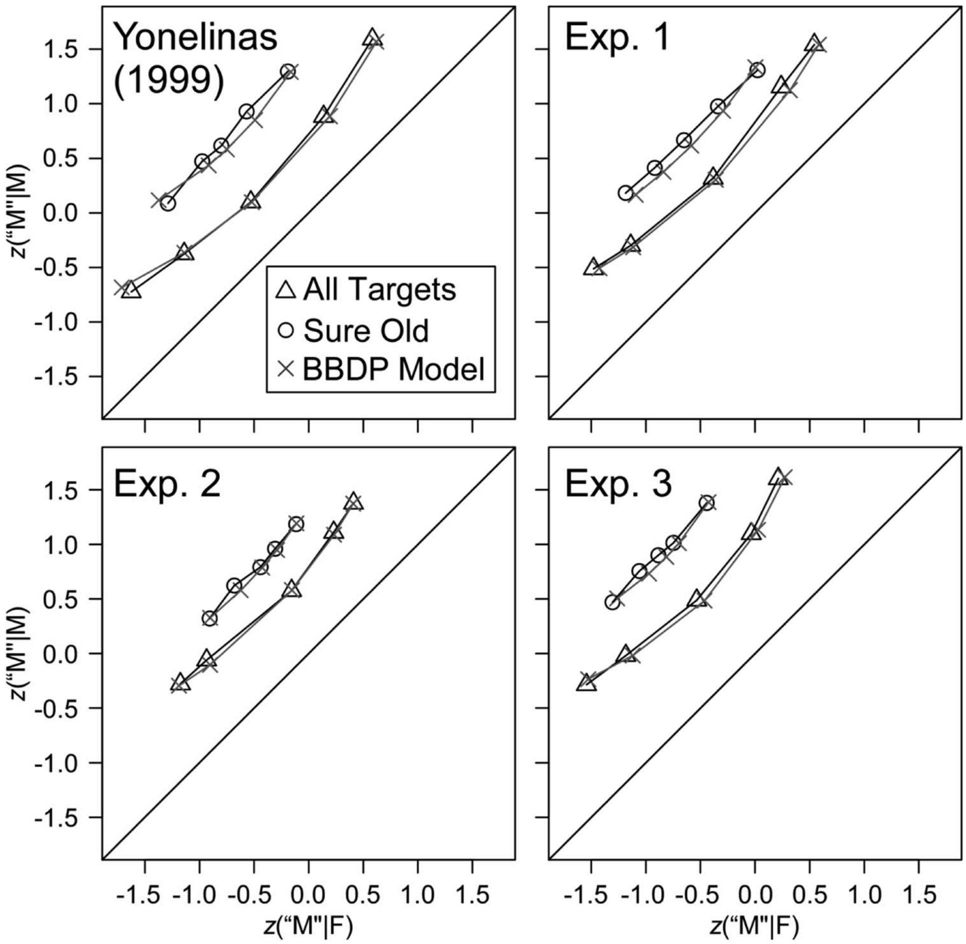

Figure 11. Fit of the bounded bivariate dual process (BBDP) model to the source *z*-transformed receiver operating characteristic (zROC) functions for all target items and just for targets receiving a "Sure Old" response (i.e., the refined source function). Panels 1–4 show results from Yonelinas (1999) and Experiments 1–3, respectively. z("M"|F) = *z*-score of the proportion of female items attributed to the male voice; z("M"|M) = *z*-score of the proportion of male items attributed to the male voice.

the empirical functions were linear. One potential concern is that the empirical functions were distorted because data were pooled across participants (Pratte, Rouder, & Morey, 2010). Like previous studies testing bivariate signal-detection models, we were unable to model data at the individual-participant level (DeCarlo, 2003; Hautus et al., 2008; Klauer & Kellen, 2010; Onyper et al., 2010). Fitting the models requires estimating the frequency of responses in 36 response categories (6 item confidence levels × 6 source confidence levels), which is infeasible with the 40 or 80 observations that each subject provided per condition. Individual-level modeling will require experiments with multiple sessions of data collection from each participant, and this is an important future goal for research in this area.

Although we acknowledge that modeling group data can be problematic, group data often support the same conclusions as individual-level modeling; indeed, group data can even support more accurate conclusions when there are few observations from each participant (Cohen, Sanborn, & Shiffrin, 2008). To determine if pooling across participants distorted the data in a manner that produced misses for the dual process model, we simulated data from this model with the parameters for each participant drawn from Gaussian distributions.[5] We performed four simulations using the number of participants and number of trials per participant in each of the four data sets, and each simulation included 1000 full replications of the experiment. We compared the quadratic coefficient for the "real" function predicted by the average parameter values to the function produced by pooling data across the simulated participants. The average parameter values were selected to match the refined source zROCs predicted by the BBDP model, and the resulting quadratic coefficients for the unpooled functions were .39, .32, .41, and .54 for the four data sets, demonstrating the u-shape predicted by the model. The coefficients

[5] The standard deviations of the across-participant distributions were .1 for the recollection parameter, .25 for the familiarity d' parameter, .25 for the position of the center response criterion, and .25 for the distance between each pair of adjacent criteria.

from the zROC functions pooled across simulated participants were .35, .28, .36, and .49. So averaging did introduce a slight downward bias in the coefficients, but the functions were still clearly u-shaped in contrast to the basically linear empirical functions, which had quadratic coefficients of −.09, .02, .02, and .09. Although this simulation does not rule out any conceivable averaging artifact, it does show that averaging the u-shaped functions predicted by the dual process model does not generally produce the linear functions observed from participants. In other words, if the individual participants had u-shaped refined zROC functions as predicted by the dual process account, then the overall function should have been u-shaped as well.

Response Time (RT) Modeling

We have noted that decision biases can affect zROC characteristics even in our bivariate signal-detection models, but models that accommodate response time (RT) distributions in addition to ROC data have additional mechanisms whereby decision processes can affect zROC slope and shape (Ratcliff & Starns, 2009, 2013; Van Zandt, 2000). In particular, RT models specify more complete decision mechanisms that accommodate the accumulation of evidence over time, as opposed to a static evidence value for each item in signal-detection models. As a result, RT models include criteria for the amount of accumulated evidence needed to produce a response, and changes in these criteria can affect zROC properties that are usually attributed to memory processes in signal-detection models.

RT models sometimes support the same conclusions about memory evidence as signal-detection models; for example, both signal-detection and RT models require unequal-variance distributions to accommodate zROC slopes less than 1 (Ratcliff & Starns, 2009; Starns & Ratcliff, 2014; Starns, Ratcliff, & McKoon, 2012). However, RT models can also support interpretations that are different than signal-detection models. For example, RT modeling by Ratcliff and Starns (2013) showed that the non-linear zROC functions reported by Ratcliff et al. (1994, Experiment 5) were produced by unequal decision boundaries across confidence levels, whereas changing criteria does not affect the linearity of the function in signal detection models.

RT models are currently not available for joint recognition and source confidence ratings, and extending RT models to multiple evidence dimensions will be a substantial challenge. Although we cannot explicitly apply RT modeling, we now consider whether our central interpretations would be likely to change if RT modeling was substituted for our signal-detection modeling; that is, whether effects that we attribute to memory are likely to be driven by decision processes.

First, we interpreted the difference between source-correct and source-incorrect zROC slopes in terms of memory processes. The fact that recognition zROC functions have a lower slope for source-correct than source-incorrect items is unlikely to reflect decision processes, even in a model with RT mechanisms. Any source details remembered during the recognition decision should provide evidence that the item was studied as opposed to changing the decision standards used to evaluate evidence. We also note that the same pattern emerged even when the recognition and source decisions were segregated into separate tests, eliminating the possibility that planning for the source response affected decision processes for the recognition response (see Starns et al.'s, 2013, data in Table 1). Thus, this pattern is probably best accounted for by characteristics of memory evidence. Given the success of the BBG model, our preferred explanation of the slope difference is that source-correct items have a wider range of item strengths than source-incorrect items. In other words, this is another case in which the variance of the evidence distributions impacts the zROC slope.

Second, we interpreted the shape of the refined source zROC function as evidence against a threshold recollection process in memory. This interpretation is more vulnerable to alternative explanations in terms of decision processes. By applying a model of RT and confidence, Ratcliff and Starns (2013) showed that recognition memory zROC functions can be u-shaped for participants with decision criteria that are biased toward low confidence responses and inverted-u-shaped for participants with decision criteria that are biased toward high confidence. In this light, the linear source zROC functions for words recognized with high confidence could possibly result from memory processes that produce a u-shape coupled with decision processes that produce an inverted-u shape. This account requires that the two factors are perfectly balanced in each of the numerous experiments showing linear refined functions, which seems implausible. Testing this account rigorously would require extending the dual process approach to RT data, and this extension would not change the fact that the signal detection version of the dual process model—the version that is currently used to estimate recollection and familiarity—is inconsistent with the data.

The recognition memory zROCs from Experiment 1 had an inverted-u shape, and none of the models were able to match this aspect of the data. Although linear zROCs are very common for recognition memory, the literature does include a few reports of nonlinear functions (e.g., Ratcliff et al., 1994). Models of RT and confidence can produce inverted-u functions based on changes in the decision criteria across levels of the confidence scale, and individual participants who show inverted-u functions also have RT distributions that are consistent with these changes in criteria (Ratcliff & Starns, 2013).

Conclusions

Our results inform the ongoing debate between theorists who propose that memory decisions are based on purely continuous evidence and theorists who propose that memory decisions are based on a mixture of continuous and threshold processes (see Pazzaglia, Dube, & Rotello, 2013, for a recent and general review of the evidence). To match the lower slope for source-correct items, the dual process model had to propose a large role for recollection in source discrimination. As a result, the model predictions for the refined source zROC functions were u-shaped in contrast to the linear empirical functions. A continuous model with bounded bivariate distributions closely matched all aspects of the data and considerably out-performed the non-bounded version of this model that has been applied in previous studies. Critically, the continuous models also matched the u-shape for the overall source functions, a feature that has been cited as strong support for a threshold recollection process (Yonelinas, 1999). Overall, our results show that jointly modeling recognition and source confidence ratings can discriminate alternative models more effectively than modeling either task in isolation.

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov & F. Csaki (Eds.), *Second International Symposium on Information Theory* (pp. 267–281). Budapest, Hungary: Akademiai Kiado.

Banks, W. P. (2000). Recognition and source memory as multivariate decision processes. *Psychological Science, 11,* 267–273. doi:10.1111/1467-9280.00254

Clark, S. E., & Gronlund, S. D. (1996). Global matching models of recognition memory: How the models match the data. *Psychonomic Bulletin & Review, 3,* 37–60. doi:10.3758/BF03210740

Cohen, A. L., Sanborn, A. N., & Shiffrin, R. M. (2008). Model evaluation using grouped or individual data. *Psychonomic Bulletin & Review, 15,* 692–712. doi:10.3758/PBR.15.4.692

DeCarlo, L. T. (2003). Source monitoring and multivariate signal detection theory, with a model for selection. *Journal of Mathematical Psychology, 47,* 292–303. doi:10.1016/S0022-2496(03)00005-1

Dennis, S., & Humphreys, M. S. (2001). A context noise model of episodic word recognition. *Psychological Review, 108,* 452–478. doi:10.1037/0033-295X.108.2.452

Dougal, S., & Rotello, C. M. (2007). "Remembering" emotional words is based on response bias, not recollection. *Psychonomic Bulletin & Review, 14,* 423–429. doi:10.3758/BF03194083

Dunn, J. C. (2004). Remember–know: A matter of confidence. *Psychological Review, 111,* 524–542. doi:10.1037/0033-295X.111.2.524

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

Elfman, K. W., Parks, C. M., & Yonelinas, A. P. (2008). Testing a neurocomputational model of recollection, familiarity, and source recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34,* 752–768. doi:10.1037/0278-7393.34.4.752

Genz, A., Bretz, F., Miwa, T., Mi, X., Leisch, F., Scheipl, F., & Hothorn, T. (2013). mvtnorm: Multivariate normal and *t* distributions (R package Version 0.9-9995) [Computer software]. Retrieved from http://CRAN.R-project.org/package=mvtnorm

Glanzer, M., Hilford, A., & Kim, K. (2004). Six regularities of source recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 1176–1195. doi:10.1037/0278-7393.30.6.1176

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. doi:10.1037/0278-7393.25.2.500

Hautus, M. J., Macmillan, N. A., & Rotello, C. B. (2008). Toward a complete decision model of item and source recognition. *Psychonomic Bulletin & Review, 15,* 889–905. doi:10.3758/PBR.15.5.889

Hilford, A., Glanzer, M., Kim, K., & DeCarlo, L. T. (2002). Regularities of source recognition: ROC analysis. *Journal of Experimental Psychology: General, 131,* 494–510. doi:10.1037/0096-3445.131.4.494

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. doi:10.1037/0278-7393.27.3.701

Klauer, K. C., & Kellen, D. (2010). Toward a complete decision model of item and source recognition: A discrete-state approach. *Psychonomic Bulletin & Review, 17,* 465–478. doi:10.3758/PBR.17.4.465

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

McClelland, J. L., & Chappell, M. (1998). Familiarity breeds differentiation: A subjective-likelihood approach to the effects of experience in recognition memory. *Psychological Review, 105,* 724–760. doi:10.1037/0033-295X.105.4.734-760

Mickes, L., Johnson, E. M., & Wixted, J. T. (2010). Continuous recollection vs. unitized familiarity in associative recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36,* 843–863. doi:10.1037/a0019755

Norman, K. A., & O'Reilly, R. C. (2003). Modeling hippocampal and neocortical contributions to recognition memory: A complementary-learning-systems approach. *Psychological Review, 110,* 611–646. doi:10.1037/0033-295X.110.4.611

Onyper, S. V., Zhang, Y. X., & Howard, M. W. (2010). Some-or-none recollection: Evidence from item and source memory. *Journal of Experimental Psychology: General, 139,* 341–364. doi:10.1037/a0018926

Parks, C. M., & Yonelinas, A. P. (2007). Moving beyond pure signal-detection models: Comment on Wixted (2007). *Psychological Review, 114,* 188–201. doi:10.1037/0033-295X.114.1.188

Pazzaglia, A. M., Dube, C., & Rotello, C. M. (2013). A critical comparison of discrete-state and continuous models of recognition memory: Implications for recognition and beyond. *Psychological Bulletin, 139,* 1173–1203. doi:10.1037/a0033044

Pitt, M. A., Myung, I. J., & Zhang, S. (2002). Toward a method of selecting among computational models of cognition. *Psychological Review, 109,* 472–491. doi:10.1037/0033-295X.109.3.472

Pratte, M. S., Rouder, J. N., & Morey, R. D. (2010). Separating mnemonic process from participant and item effects in the assessment of ROC asymmetries. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36,* 224–232. doi:10.1037/a0017682

Ratcliff, R., McKoon, G., & Tindall, M. H. (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. doi:10.1037/0278-7393.20.4.763

Ratcliff, R., Sheu, C.-F., & Gronlund, S. (1992). Testing global memory models using ROC curves. *Psychological Review, 99,* 518–535. doi:10.1037/0033-295X.99.3.518

Ratcliff, R., & Starns, J. J. (2009). Modeling confidence and response time in recognition memory. *Psychological Review, 116,* 59–83. doi:10.1037/a0014086

Ratcliff, R., & Starns, J. J. (2013). Modeling confidence judgments, response times, and multiple choices in decision making: Recognition memory and motion discrimination. *Psychological Review, 120,* 697–719. doi:10.1037/a0033152

R Development Core Team. (2010). R: A language and environment for statistical computing [Computer software]. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics, 6,* 461–464. doi:10.1214/aos/1176344136

Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: REM—Retrieving effectively from memory. *Psychonomic Bulletin & Review, 4,* 145–166. doi:10.3758/BF03209391

Shimamura, A. P., & Wickens, T. D. (2009). Superadditive memory strength for item and source recognition: The role of hierarchical relational binding in the medial temporal lobe. *Psychological Review, 116,* 1–19. doi:10.1037/a0014500

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. doi:10.3758/BF03195305

Starns, J. J., Hicks, J. L., Brown, N. L., & Martin, B. A. (2008). Source memory for unrecognized items: Predictions from multivariate signal detection theory. *Memory & Cognition, 36,* 1–8. doi:10.3758/MC.36.1.1

Starns, J. J., Pazzaglia, A. M., Rotello, C. M., Hautus, M. J., & Macmillan, N. A. (2013). Unequal-strength source zROC slopes reflect criteria placement and not (necessarily) memory processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39,* 1377–1392. doi:10.1037/a0032328

Starns, J. J., & Ratcliff, R. (2014). Validating the unequal-variance assumption in recognition memory using response time distributions instead of ROC functions: A diffusion model analysis. *Journal of Memory and Language, 70,* 36–52.

Starns, J. J., Ratcliff, R., & McKoon, G. (2012). Evaluating the unequal-variance and dual-process explanations of zROC slopes with response time data and the diffusion model. *Cognitive Psychology, 64,* 1–34. doi:10.1016/j.cogpsych.2011.10.002

Van Zandt, T. (2000). ROC curves and confidence judgments in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 582–600. doi:10.1037/0278-7393.26.3.582

Wixted, J. T. (2007). Dual-process theory and signal-detection theory of recognition memory. *Psychological Review, 114,* 152–176. doi:10.1037/0033-295X.114.1.152

Wixted, J. T., & Mickes, L. (2010). A continuous dual-process model of remember/know judgments. *Psychological Review, 117,* 1025–1054. doi:10.1037/a0020874

Wixted, J. T., & Squire, L. R. (2010). The role of the human hippocampus in familiarity-based and recollection-based recognition memory. *Behavioural Brain Research, 215,* 197–208. doi:10.1016/j.bbr.2010.04.020

Yonelinas, A. P. (1994). Receiver-operating characteristics in recognition memory: Evidence for a dual-process model. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 1341–1354. doi:10.1037/0278-7393.20.6.1341

Yonelinas, A. P. (1999). The contribution of recollection and familiarity to recognition and source-memory judgments: A formal dual-process model and an analysis of receiver operating characteristics. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 1415–1434. doi:10.1037/0278-7393.25.6.1415

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. doi:10.1006/jmla.2002.2864

Appendix

Formal Descriptions of the Bivariate Gaussian (BG), Bounded Bivariate Gaussian (BBG), Bivariate Dual-Process (BDP), and Bounded Bivariate Dual Process (BBDP) Models

BG Model

The memory distribution for each item type is defined by parameters for the mean evidence value in the item (μ_I) and source (μ_S) dimensions, the standard deviations in the item (σ_I) and source (σ_S) dimensions, and the correlation between item and source evidence (ρ). For all models, new items have a mean of zero and a standard deviation of one in both dimensions. All of the models also include five criteria, λ_{I1}–λ_{I5}, to map item evidence values onto responses on the recognition confidence scale, which ranged from 1 (high confidence "new") to 6 (high confidence "old"). At each level of item confidence, there were separate parameters for the source confidence criteria (λ_{S1}–λ_{S5}) to map source evidence values onto confidence ratings from 1 (high confidence "Source B") to 6 (high confidence "Source A"). In all of the reported fits, the source criteria were constrained to be equal for all three levels of confidence that the item was new (i.e., item confidence levels 1–3).

For a bivariate normal distribution, the probability density for a pair of strength values (x, y) is equal to the marginal probability density of the y value multiplied by the conditional probability density of the x value given y. The marginal distribution for item evidence (y) is a univariate normal distribution with mean μ_I and standard deviation σ_I. Equations A1 and A2 give the mean and standard deviation for the conditional distribution of source evidence (x).

$$\mu_{S|I} = \mu_S + \frac{\rho\sigma_S(y - \mu_I)}{\sigma_I} \tag{A1}$$

$$\sigma_{S|I} = \sqrt{(1-\rho^2)\sigma_S^2} \tag{A2}$$

Therefore, the probability density function for the bivariate Gaussian can be written as

$$f(x, y) = \left[\frac{1}{\sigma_{S|I}}\phi\left(\frac{x - \mu_{S|I}}{\sigma_{S|I}}\right)\right]\left[\frac{1}{\sigma_I}\phi\left(\frac{y - \mu_I}{\sigma_I}\right)\right], \tag{A3}$$

where ϕ is the density function for a standard univariate Gaussian distribution. The probability density for a value on a Gaussian distribution with any mean and standard deviation can be found by converting the value to a z-score, taking the probability density of this z-score on a standard Gaussian distribution ($\mu = 0$ and $\sigma = 1$), and dividing by the standard deviation of the desired distribution. Thus, the term in the first set of brackets gives the conditional density of a value of source evidence (x) given the value of item evidence (y), and the second term gives the marginal density of the item evidence value. Multiplying the two terms gives the bivariate density.

(*Appendix continues*)

We derived predictions for the BG model using the mvtnorm package (Genz et al., 2013) in (R R Development Core Team, 2010). The "pmvnorm" function in this package returns integrals of the multivariate normal density function between two cutoff values on the x and y dimensions. For the bivariate memory model, this function takes four arguments: (1) a vector of lower cutoffs in the source and item dimensions; (2) a vector of upper cutoffs in the source and item dimensions; (3) a vector of distribution means in the source and item dimensions (μ_S and μ_I, respectively); and (4) a covariance matrix in which the value in the first row and first column is the variance on the source dimension (σ_S^2), the value in the second row and second column is the variance on the item dimension (σ_I^2), and the values on the diagonal are the covariance ($\rho \times \sigma_S \times \sigma_I$).

To compute the proportion of responses in each item and source rating category, we created vectors of cutoff values to use in the pmvnorm function. For each vector, the first element had a value of negative infinity, the next five elements were the five criteria from lowest to highest, and the last element had a value of infinity. The first and last elements accommodated the proportion of responses below the lowest criterion and above the highest criterion, respectively. We used a single vector for cutoffs on the item dimension (y_1–y_7). On the source dimension, we used a separate vector (x_1–x_7) for each level of item confidence; for example, $x_3|2$ indexes the third cutoff value on the source dimension given that the item rating was 2. With vectors constructed in this way, the joint probability of an item rating j and a source rating k is found using lower bounds $x_k|j$ and y_j and upper bounds $x_{k+1}|j$ and y_{j+1}. Thus, the prediction equation for the BG model is

$$p(j,k) = \int_{x_k|j}^{x_{k+1}|j} \int_{y_j}^{y_{j+1}} \left[\frac{1}{\sigma_{S|I}}\phi\left(\frac{x-\mu_{S|I}}{\sigma_{S|I}}\right)\right]\left[\frac{1}{\sigma_I}\phi\left(\frac{y-\mu_I}{\sigma_I}\right)\right]dydx, \tag{A4}$$

where $\mu_{S|I}$ and $\sigma_{S|I}$ are given by Equations A1 and A2, respectively.

BBG Model

This model has the same parameters as the BG model (although some of the parameters have different interpretations as described in the next section). Ensuring that the source distributions do not cross requires a slight change to the equation for the conditional source mean at a given item evidence value,

$$\mu_{S|I} = \begin{cases} max\left[\mu_S + \dfrac{\rho\sigma_S(y-\mu_I)}{\sigma_I}, 0\right] & for \quad Source \quad A \\ min\left[\mu_S + \dfrac{\rho\sigma_S(y-\mu_I)}{\sigma_I}, 0\right] & for \quad Source \quad B, \end{cases} \tag{A5}$$

such that the conditional source mean never crosses to the opposite side of the source continuum as the overall source mean. Equation A2 still describes the conditional standard deviation in source evidence. Substituting the revised formula for $\mu_S|I$ into Equations A3 and A4 gives the probability density function and the prediction equation for the bounded bivariate Gaussian distribution, respectively.

To get predictions for the BBG model, we first found the point on the item axis where the conditional source mean was zero. To do so, we entered a value of zero for $\mu_{S|I}$ in Equation A1 and solved for y. This produced the formula

$$cp = \mu_I - \frac{\mu_S\sigma_I}{\rho\sigma_S}, \tag{A6}$$

where cp stands for *crossing point*—that is, the y value where $\mu_{S|I}$ equals zero. We again used the mvtnorm function to find the predicted response proportions as described above, but this requires using different distributional parameters based on whether the cutoffs defining the region to integrate are below or above the crossing point on the item dimension. Above the crossing point, the bounded distribution is identical to a standard bivariate normal distribution with the same parameters. Thus, we used the same distributional parameters as described for the BG model when both of the item evidence cutoff values were above the crossing point. Below the crossing point, the bounded distribution has the same item mean and standard deviation as a standard bivariate Gaussian, but the source mean is zero, the source standard deviation is $\sigma_{S|I}$ (Equation A2), and the correlation between source and item evidence is zero. When both of the item evidence cutoff values were below the crossover point, we defined predictions with the mean vector and covariance matrix adjusted to the appropriate values. When the crossover point fell between the two item evidence cutoff values, we made separate function calls for the region below and above the crossover (using the appropriate parameter values) and added them together.

Parameter Interpretation in the BBG Model

The memory evidence parameters that we report for the BBG model characterize the distributions *before* the correction is applied to eliminate below-chance source performance at low item strengths. This facilitates comparisons with the BG model: If the parameters of the two models are the same, then the evidence distributions will be the same above the cutoff point where source performance crosses over to reverse discriminability. However, the reader should keep in mind that some of the parameters have slightly different interpretations in the two models. First, the correlation parameter in the BBG model only describes the correlation between item and source evidence above the cutoff point. Below the cutoff, the mean source evidence is zero for every value of item evidence, so the correlation is zero as well. Second, applying the correction alters the overall source evidence mean in

(*Appendix continues*)

the BBG model, so the marginal mean is no longer equal to μ_S. Instead, the μ_S parameter is the conditional mean of the source distribution at the mean value of item evidence. For the BG model, the μ_S parameter is both the overall source mean and the conditional source mean at the item mean. Third, applying the correction also alters the overall variability of the source distribution such that the standard deviation does not equal σ_S. However, if the two models have the same σ_S parameter, then the *conditional* variability in source evidence at a given level of item evidence is equal for the two models.

BDP Model

The familiarity process in this model matches the BG model discussed above with σ_I and σ_S equal to 1. In addition to the parameters already described for the BG model, the R_I parameter gives the proportion of items that are successfully recollected as being presented on the study list, and the R_S parameter gives the proportion of times that source details are part of the recollected information. R_I and R_S are always zero for lure items. We divided the prediction function into three separate components, one for

Lures

		Sure "B" 1	2	3	4	5	Sure "A" 6
Sure "New"	1	*p*1	*p*1	*p*1	*p*1	*p*1	*p*1
	2	*p*1	*p*1	*p*1	*p*1	*p*1	*p*1
	3	*p*1	*p*1	*p*1	*p*1	*p*1	*p*1
	4	*p*1	*p*1	*p*1	*p*1	*p*1	*p*1
	5	*p*1	*p*1	*p*1	*p*1	*p*1	*p*1
Sure "Old"	6	*p*1	*p*1	*p*1	*p*1	*p*1	*p*1

Source A Items

		Sure "B" 1	2	3	4	5	Sure "A" 6
Sure "New"	1	*p*1	*p*1	*p*1	*p*1	*p*1	*p*1
	2	*p*1	*p*1	*p*1	*p*1	*p*1	*p*1
	3	*p*1	*p*1	*p*1	*p*1	*p*1	*p*1
	4	*p*1	*p*1	*p*1	*p*1	*p*1	*p*1
	5	*p*1	*p*1	*p*1	*p*1	*p*1	*p*1
Sure "Old"	6	*p*1 + *p*2	*p*1 + *p*2	*p*1 + *p*2	*p*1 + *p*2	*p*1 + *p*2	*p*1 + *p*2 + *p*3

Source B Items

		Sure "B" 1	2	3	4	5	Sure "A" 6
Sure "New"	1	*p*1	*p*1	*p*1	*p*1	*p*1	*p*1
	2	*p*1	*p*1	*p*1	*p*1	*p*1	*p*1
	3	*p*1	*p*1	*p*1	*p*1	*p*1	*p*1
	4	*p*1	*p*1	*p*1	*p*1	*p*1	*p*1
	5	*p*1	*p*1	*p*1	*p*1	*p*1	*p*1
Sure "Old"	6	*p*1 + *p*2 + *p*3	*p*1 + *p*2	*p*1 + *p*2	*p*1 + *p*2	*p*1 + *p*2	*p*1 + *p*2

Figure A1. Schematic showing how the three components of the prediction function for the bounded dual process and bounded bivariate dual process models combine to determine the proportion of responses in each cell of the response matrix. The *p*1 component represents trials in which both the item and recognition responses are based on familiarity, the *p*2 component represents trials in which the item response is based on recollection and the source response is based on familiarity, and the *p*3 component represents trials in which both the item and source responses are based on recollection.

(*Appendix continues*)

cases in which both the item and source rating are based on familiarity ($p1$), one for cases in which the item rating is based on recollection but the source rating is based on familiarity ($p2$), and one for cases in which both the item and source ratings are based on recollection ($p3$). The proportion of times that both decisions are based on familiarity leading to an item rating j and a source rating k is

$$p1(j,k) = (1 - R_I)\int_{x_k|j}^{x_{k+1}|j}\int_{y_j}^{y_{j+1}}\left[\frac{1}{\sigma_{S|I}}\phi\left(\frac{x-\mu_{S|I}}{\sigma_{S|I}}\right)\right] \times\left[\frac{1}{\sigma_I}\phi\left(\frac{y-\mu_I}{\sigma_I}\right)\right]dydx, \quad \text{(A7)}$$

which is simply the joint probability that recollection will fail (1 – R_I) and the item and source familiarity values will fall between the criteria defining the response category. The proportion of times that non-source recollection succeeds to produce an item rating of 6 (high-confidence "old") and source familiarity supports a source rating of k is

$$p2(1,k) = R_I(1 - R_S)\int_{x_k|1}^{x_{k+1}|1}\int_{-\infty}^{\infty}\left[\frac{1}{\sigma_{S|I}}\phi\left(\frac{x-\mu_{S|I}}{\sigma_{S|I}}\right)\right] \times\left[\frac{1}{\sigma_I}\phi\left(\frac{y-\mu_I}{\sigma_I}\right)\right]dydx, \quad \text{(A8)}$$

which is the joint probability that item recollection will succeed (R_I), source recollection will fail (1 – R_S), and the source familiarity value will fall between the source criteria defining the response category. The source ratings are based on the marginal distribution of source familiarity across all item evidence values. This follows from the assumption that recollection and familiarity are independent, so items with successful recollection come from the full spectrum of item familiarity values. The proportion of times that recollection will succeed and the recollected information will include source details is

$$p3 = R_I R_S, \quad \text{(A9)}$$

where the j and k subscripts are not needed because the item response is always 6 (high-confidence old) and the source response is always high-confidence correct (1 for Source B items, 6 for Source A items).

Figure A1 shows how the overall proportion of responses (p) is calculated from the three components for each item type and rating category. Recollection never succeeds for new items, so all of the predictions for these items are based only on component $p1$. For target items, recollection always leads to a high-confidence "old" response, so all categories with an item rating of 1–5 are also based only on component $p1$. When recollection succeeds but is not source specific, any of the source ratings can be made based on the level of source familiarity. Thus, component $p2$ affects all categories with an item rating of 6 regardless of the source rating. When recollection succeeds and the recollected information includes source details, the item rating is always 6 and the source rating is always high-confidence correct. Thus, component $p3$ affects only the item–source combination 6–6 for Source A items and 6–1 for Source B items.

BBDP Model

The prediction equation for the BBDP model is the same as the equation for the BDP model, except that $\mu_{S|I}$ is calculated with Equation A5 instead of Equation A1.

Received August 7, 2013
Revision received January 13, 2014
Accepted February 3, 2014 ■