

**A multidimensional analysis of learning and forgetting in recognition memory**

Michael Diaz and Aaron S. Benjamin

University of Illinois, Urbana-Champaign

Corresponding author:

Michael Diaz

Department of Psychology

Washington University In St. Louis

1 Brookings Drive

St. Louis, MO 63130

mdiaz@wustl.edu

## Abstract

This paper applies a bivariate continuous-strength model to understanding the relationship between recognition and source memory, and to evaluating whether learning and forgetting in those tasks follows the same or different information pathways. The model is grounded in the source monitoring framework, which suggests that the strength of a test probe depends upon what aspects of a memory representation are queried. The fits of the model suggest that that the effects of study time (Experiment 1) are not global across the memory trace—that is, that performance on source memory and recognition tests do not increase proportionately with increased learning—but that the effects of forgetting (Experiment 2) are. The results further support the viability of continuous-strength models in interpreting performance on memory-judgment tasks and suggest that forgetting does not induce a reversion of the processes that operate during encoding.

The goal of this article is to apply multivariate signal-detection theory (SDT, Ashby, 1992; DeCarlo, 2003, Hautus, Macmillan & Rotello, 2008) and the source monitoring framework (Johnson, Hashtroudi, & Lindsay, 1993) to understanding the effects of learning and forgetting on recognition, and to evaluate the performance of that model. Along the way, the performance of the multivariate model is assessed with respect to restricted versions of that model, as well as to dual-process signal-detection theory (DPSDT; Yonelinas, 1994, 1999), which provides an alternative theoretical conception and is a popular benchmark in the field. Before we consider the relationship between learning and forgetting, we review the nature and origin of these competing theories and provide the experimental and analytic landscape on which our analysis of learning and forgetting will take place.

In the simplest case, signal detection theory explains how a stimulus is assigned to one of two mutually exclusive and exhaustive statistical hypotheses (e.g., “old” or “new”; Egan, 1958; Green & Swets, 1966), each with its own unique probability distribution (Swets, 1963). In the case of recognition, there is some memorial distribution for the studied items that is distinct from the distribution of unstudied items. For any given stimulus, a participant’s task can be said to be to infer which of those distributions that stimulus belongs to. It is this perspective that underlies all of the theories presented and evaluated in this article. We will not consider alternative approaches here, such as theories with probabilistic responding (Parks, 1966) or noisy decision-making (Benjamin, Diaz, & Wee, 2009).

### Strength theories of recognition

Strength theories of recognition tie the decision apparatus of signal detection theory to a mnemonic process by suggesting that both the old and new distributions are continuous functions of some unitary variable “strength.” Process models commonly describe this strength variable as the outcome of comparing a test stimulus to the contents of memory as a whole (Hintzman, 1986; Murdock, 1982; Shiffrin & Steyvers, 1997). There is naturally noise in such a process, but on average, previously studied stimuli yield a higher “degree of match,” or strength, than the non-studied stimuli. A recognition decision can then be made by choosing an appropriate decision criterion and calling any stimulus whose strength exceeds that criterion old. This strategy will result in a greater number of truly old items than new items being called old.

Yonelinas (1999) argued that, when two classes of items are studied under equivalent conditions, and thus equal in strength, strength theories have no way of distinguishing between those items. One experimental task that reveals this ability is a source memory task. In a source memory task, people study stimuli that appear in various different contexts (usually two). At test, they are given a stimulus and have to indicate which of the contexts that stimulus was tested in. To the extent that the study was well controlled—i.e., the different contexts are equally salient and equally memorable—items from each of the contexts will, on average, have equal memory strength (Yonelinas, 1999). As such, there is no criterion participants can use on a pure strength axis that would lead to a differential rate of responding to the different stimulus types. In other words, no matter where the criterion is set, the rate of providing a given response should be the same for all stimulus types. This would imply that people should

not be able to perform above chance on source memory tests. The fact that they can is evidence against strength theory.

An alternative view to the single-process strength view is the dual-process view. According to this view, recognition memory relies on two distinct processes instead of just a single strength variable (Atkinson, & Juola, 1973, 1974; Jacoby, 1991; Mandler, 1980; Yonelinas, 1994, 1997, 1999, 2002). The first process, frequently called familiarity, is essentially the same type of strength variable evident in the single-process models. The second process, however, is qualitatively different. This process, usually termed recollection, refers to retrieval of the episode in which a stimulus was studied. When a studied stimulus is recollected, the participant has access to the qualitative details of the study episode. In such cases, the participant can infer with high confidence that the stimulus is old since they remember the details of when it was studied.

In the typical source memory task, dual-process theorists originally posited that familiarity is, on average, the same for equivalently studied items from different sources and thus useless in distinguishing between them (Yonelinas, 1999). Any above-chance performance therefore depends solely upon the process of recollection<sup>1</sup>. When a studied item is recollected, the participant is able to remember details of the source and uses that information to make correct source memory response. When a studied item is not recollected, the participant essentially responds at random (i.e., based on familiarity). This model of source memory performance implies that source memory relies entirely upon recollection, and makes predictions that will be laid out in more detail below.

---

<sup>1</sup> More recent versions of dual-process theory have softened this claim (since the empirical data has not supported it). Familiarity can now support source judgments when item and source information are “unitized.” Unitization is discussed further below.

### *Source Monitoring Framework: Strength Theory Revisited*

The problem with applying strength theories to source memory, as originally postulated by dual-process theories, is that the different types of old items are all equal in memory strength. How can two equally studied, equally salient, items yield different strengths or “goodness of match” on a continuous variable? It is this failing of strength theories that motivate dual-process approaches, but a reconciliation may be possible. In particular, the answer may be found in the Source Monitoring Framework proposed by Johnson et al. (1993). According to the source monitoring framework, when people search their memories for a source, they tend to look in the areas of memory that can confirm or disconfirm the presence of that source. For example, when trying to remember whether a particular stimulus was previously seen, people will assess how much evidence there is for that hypothesis in visual areas of memory. Lack of evidence in auditory areas may even be taken as additional evidence in favor of visual presentation.

The source monitoring framework implies that the “degree of match” or strength of an item can be different depending on what the test question is. Marsh and Hicks (1998) found support for such a view by using a variant of a source memory test called an exclusion test (Jacoby, 1991). In an exclusion test, participants are asked whether a test item occurred in a particular source. The participant is to respond “yes” if it was studied in that source, and “no” if it was studied in a different source or if it was not studied at all. Marsh and Hicks (1998) had participants read some words and generate other words from anagrams during study. They found that participants were better able to discriminate between these two types of studied words when they were asked “was it

generated?” than when they were asked “was it read?” Marsh and Hicks (1998) suggested that the prior question directed participants’ attention towards (the memory of) the processing performed during study, which was more useful in distinguishing between the different item types, and the latter question directed their attention more towards visual or semantic aspects of the memory trace, which was less useful in distinguishing the item types.

In a converging line of evidence, Jacoby, Shimizu, Daniels, and Rhodes (2005) found evidence that participants process the novel items during a recognition test differently depending on what they are being tested on. Specifically, they had participants process some items deeply and some item shallowly during a study phase. They then were given three different recognition tests. In the first test, participants were correctly informed that all the old items were from the same levels-of-processing condition in the study phase (i.e., they were all either from the deep or from the shallow condition). In the second test, they were correctly informed that all the old items were from the other levels-of-processing condition. Finally, on the critical third test, participants were tested on the novel items that had appeared on the first two tests. Jacoby et al. (2005) found that participants better remembered novel items that had previously appeared on the “deep” test than those that appeared on the “shallow” test. They interpreted these findings as suggesting that participants were processing the items on these tests differently, and presumably in a manner consistent with how the old items had originally been processed. That is, when participants were tested on “deep” items, they searched “deep” areas of memory for evidence of a test item, and this process yielded superior memory for the novel items that were also processed deeply.

Jacoby et al.'s (2005) results are complimented by findings from studies using event-related potentials (ERPs). ERPs elicited by novel items during “deep” tests are different than those elicited by novel items during “shallow” tests (Rugg, Allan, & Birch, 2000). Similarly, ERPs for novel items are different on recognition tests than they are on source memory tests (Ranganath & Paller, 2000). Source memory tests in general tend to elicit more frontal positivity than recognition tests, even for the exact same stimuli (Senkfor & Van Patten, 1998; Van Patten, Senkfor, & Newberg, 2000). These findings are consistent with the view that different tests query memory in qualitatively different ways.

#### Competing theories and the receiver operating characteristic (ROC)

The domain in which these theories have been pitted against one another is in predicting the form of the Receiver Operating Characteristic (ROC). An ROC is a plot of the hit rate against the false alarm rate across all possible response criteria. With a simple yes-no response, there is only one hit–false-alarm pair generated for any given participant. In contrast, a multiple-rating response scale yields multiple hit–false-alarm pairs that can be useful in distinguishing the predictions of different theories.

Strength theories frequently assume that the old and new distributions are normally distributed (see Figure 1a). This assumption makes specific predictions about the shape of the ROC. For an infinitely conservative criterion, every old and new item should be rejected. This would lead to a hit–false-alarm pair of (0,0). In contrast, an infinitely lenient criterion would lead to a hit–false-alarm pair of (1,1). A plot of all possible criteria between these two extremes yields a monotonic curvilinear function such as the ones shown in Figure 2a. The solid line in Figure 2a shows an ROC generated when the old and new distributions are of equal variance. This ROC is symmetric with



respect to the minor diagonal. The dashed line shows an ROC generated when the signal distribution has 25% greater variance than the new distribution. This latter model, called the Unequal Variance Signal Detection (UVSD) model, is what is typically supported when strength models are fit to recognition data (Glanzer, Kim, Hilford, & Adams, 1999; Heathcote, 2003; Matzen & Benjamin, 2009; Ratcliff, Sheu, & Gronlund, 1992; Yonelinas, 1994; Wixted, 2007). The degree to which these curves depart from the chance diagonal is related to the distance between the means of the old and new distributions in Figure 1a. Figure 2b shows this plot when the axes are transformed to z-scores using the inverse of the cumulative normal distribution. Strength theories predict that this plot (the zROC) will be a straight line with slope equal to the ratio of the standard deviations from the new and old distributions, respectively (Macmillan & Creelman, 2005).

Dual-process theories make somewhat different predictions about the shape of the recognition memory ROC. The first dual-process model that was proposed that predicted a particular ROC curve was the Dual-Process Signal-Detection (DPSD) model (Yonelinas, 1994). This model suggests that recollection is an all-or-none threshold process. That is, an old item is either recollected or it is not. If the item is recollected, then participants are certain that it is old and respond with the highest confidence old response. If an item is not recollected then participants respond according to the item's familiarity, which is thought to be an equal-variance signal-detection process. New items are never recollected in this model (i.e., there is no illusory recollection) and as such all responses to new items are based on the equal-variance familiarity process. Specifically, the probability of a response above the  $j^{\text{th}}$  criterion,  $c_j$  is:

$$P(X > c_j | \text{old}) = R + (1-R) * \Phi(d - c_j) \quad 1$$

$$P(X > c_j | \text{new}) = \Phi(-c_j) \quad 2$$

where  $\underline{X}$  is the response variable,  $\underline{R}$  is the proportion of old items that are recollected,  $\underline{d}$  is the distance between the old and new distributions, and  $\Phi$  is the cumulative normal distribution function. The shape of the ROC curve predicted by the DPSD model is shown in Figure 2c. For an infinitely conservative criterion the model predicts a hit–false-alarm pair of  $(0, \underline{R})^2$ . In other words, the model predicts that  $100*\underline{R}$  percent of the old items will always be endorsed no matter how strict the criterion. The shape of the zROC function predicted by the DPSD model is shown in Figure 2d. The zROC departs from linearity for the most conservative points. Moreover, this departure from linearity increases as  $\underline{R}$  increases. When  $\underline{R}=0$ , the model reduces to an equal-variance signal detection model with a linear zROC with a slope of 1. As  $\underline{R}$  increases, so does the curvilinearity of the zROC. More recently, models have been developed that relax the all-or-none threshold assumption and allow recollection at variable degrees of confidence (Kelley & Wixted, 2001; Macho, 2002, 2004; Sherman, Atri, Hasselmo, Stern, & Howard, 2003; Yonelinas & Parks, 2007; see also DeCarlo, 2002, 2003a for a mathematically equivalent strength model). Such models can produce close approximations to all four panels in Figure 2.

The empirical evidence from recognition ROCs has come down mostly in favor of the UVSD (or the related Variable Recollection Dual Process model, which is indistinguishable from it under some conditions). The zROC has been consistently found to be linear for recognition with a slope of about 0.8 (Glanzer, Kim, Hilford, & Adams,

---

<sup>2</sup> There is no real value that you can assign to  $c_j$  in Equation 1 that would yield a hit rate lower than  $R$ .

1999; Healy, Light, and Chung, 2005; Heathcote, 2003; Ratcliff, McKoon, & Tindall, 1994; Ratcliff et al., 1992; see Wixted, 2007, for a review). The UVSD always predicts linear zROCs and the parameter that accounts for the slope is the ratio of the standard deviations of the underlying distributions. The DPSD model, on the other hand, only predicts linear zROCs when  $\underline{R}$  is zero, but only predicts slopes less than 1 when  $\underline{R} > 0$ . That is, as pointed out by Heathcote (2003), the parameter in the DPSD model that creates slopes less than 1 is the very same parameter that makes the zROC curvilinear. Thus, as the slope of the zROC decreases, the curvilinearity should increase. There is no evidence of such a relationship in item recognition (Glanzer et al., 1999; Heathcote, 2003). Recognition zROCs tend to be linear even when the slope is very shallow.

Dual-process model of source memory. Yonelinas (1999) developed the DPSD model to be able to predict the ROC functions for source memory tasks, and it is in this domain in which they have greater success. Specifically, the probability of a response above the  $j^{\text{th}}$  criterion,  $c_j$  is:

$$P(X > c_j | \text{source A}) = R + (1-R) * \Phi(d - c_j) \quad 5$$

$$P(X > c_j | \text{source B}) = (1-R) * \Phi(d - c_j), \quad 6$$

where the scale is arbitrarily defined as higher values being more confidence that the item was from source A (reversing the scale would merely swap the “given” portions of equations 5 and 6). Equation 5 is analogous to “hits” (calling a source A item “source A”) and Equation 6 is analogous to false alarms (calling a source B item “source A”). If a “target” (a source A item) is recollected it is always endorsed (i.e., it is called “source A”). Conversely, if a “lure” (a source B item) is recollected it is never endorsed (i.e., it is *never* called “source A”). All the non-recollected items are endorsed according to the

strength component. However, since the target and lure distributions both have the same mean,  $\underline{d}$ , the rate of endorsing unrecalled items is the same. This model implies that above-chance performance on source memory depends entirely on recollection. All responses based on familiarity will be at chance levels. Moreover, it predicts that the source memory ROC will be a straight line going from the point  $(0, \underline{R})$  to the point  $(1 - \underline{R}, 1)$ . That is,  $100 * \underline{R}$  percent of the targets will always be endorsed,  $100 * \underline{R}$  percent of the lures will *never* elicit false alarms, and all criteria from  $-\infty$  to  $+\infty$  will yield the exact same endorsement rate for non-recalled items from both item types (i.e., a linear ROC with slope of 1). Yonelinas (1999) fit this model to three experiments using a source memory task and found that the ROC for all three were fit well by a linear function, as predicted by DPSD. According to Wixted (2007), this result was a major reason for the growth in popularity of the DPSD model. More recently, dual-process theorists have softened the assumption that familiarity is not useful for source memory by arguing that source information can be “unitized” with item information, such that familiarity can distinguish between different sources (Haskins, Yonelinas, Quamme, & Ranganath, 2008). Such a view no longer predicts linear ROCs (though the intercepts remain the same), but also damages the central argument as to why source memory cannot be based on a strength process. We do not further pursue the reach of dual-process unitization models, though we do note that the empirical evaluation of this theory using ROC analysis has not been supportive of the unitization claim (Mickes, Johnson, & Wixted, 2010).

Many studies have since found curvilinear source memory ROCs, making Yonelinas’s (1999) results the exception rather than the rule (Glanzer, Hilford, & Kim,

2004; Hilford, Glanzer, Kim, & DeCarlo, 2002; Onyper, Zhang, & Howard, 2010; Qin, Raye, Johnson, & Mitchell, 2001; Slotnick & Dodson, 2005; Slotnick, Klein, Dodson, & Shimamura, 2000; see Wixted, 2007, for a review). In a particularly telling demonstration, Slotnick and Dodson (2005) reanalyzed the data from Yonelinas (1999). According to DPSD theory, when an item is recollected in recognition memory it is given the highest confidence response. According to Yonelinas (1999), it was this same recollection process that created the above-chance linear ROCs for source memory. Slotnick and Dodson (2005) looked at one experiment from Yonelinas (1999) in which participants made both a recognition decision and a source memory decision for every old item. Looking at the source memory ROCs for every old item (regardless of how they rated it in the recognition task) yielded the same linear ROCs that Yonelinas (1999) found. However, when they constrained the set of items to only those that were given the highest rating on the recognition task, the linearity disappeared and the ROCs were highly curvilinear. According to DPSD, all items that were recollected would have been included in this set, since recollected items are given the highest confidence response.

#### Strength theory of recognition and source memory: A unified conceptualization

The relation between different memory tests can be estimated directly by a multivariate extension of the UVSD model (Banks, 2000; Wickens, 1992; DeCarlo, 2003b; Hautus et al., 2008). In this model, each memory test is placed on an axis orthogonal to each of the other memory tests. Participants make a decision for a given test by placing criteria perpendicular to that test's axis and endorsing items whose strength falls on one side of these criteria. Figure 3a gives an example of what a bivariate signal detection model may look like for recognition and source memory tests (DeCarlo,

2003b; Hilford, Glanzer, Kim, & DeCarlo, 2002; Hautus et al., 2008; Slotnick & Dodson, 2005). The bottom-middle distribution represents the new distribution. The top-left and top-right distributions represent two old distributions studied in two different sources (arbitrarily called source 1 and source 2, respectively). The dimension going back “in depth” in Figure 3a represents the recognition memory dimension. If we were to take the integral of the bivariate curves along the other dimension, we would get a standard univariate signal detection model for recognition memory. That is, both of the old distributions would have the same mean, and it would be greater than the mean of the new distribution. Both old distributions would also have the same variance, which is slightly greater than the variance of the new distribution. According to this model, the memory strength obtained from a recognition task would not be able to distinguish between the two different contexts. In contrast to this, the horizontal dimension in Figure 3a, labeled source memory, would be able to distinguish between the different contexts. Taking the integral with respect to the recognition dimension would yield a univariate signal detection model for source memory. In this particular example, the new distribution would have a mean equidistant between the means of the two old distributions.

An easier way to look at this model is to look at a two-dimensional iso-likelihood contour plot instead of a three-dimensional bivariate plot (Ashby, 1992). An iso-likelihood contour plot can be created by passing a plane of constant probability (i.e., height) through the three-dimensional plot, as illustrated in Figure 3b. The set of points on the plane that intercept a given bivariate curve form an ellipse. If the bivariate curve has zero covariance, then the major and minor axes of the ellipse will be parallel to and

perpendicular to the coordinate axes. If the covariance is zero *and* the variances on both dimensions are equal, then the ellipse will be a circle. Figure 3c shows the iso-likelihood contours for the bivariate signal detection model in Figure 3a and 3b. It is evident from the major axes of these ellipses that both old distributions have covariances such that items that have higher strength on the recognition dimension also tend to have a source strength that is further away from the new distribution. As we will review below, such models have had success in fitting recognition and source memory performance jointly, and have outperformed DPSD models.

*Test covariance and global effects of learning.*

DeCarlo (2003b) and Hautus et al. (2008) fit a bivariate signal detection model and found covariances for the old distributions that were qualitatively similar to those in Figure 3a-c. That recognition and source strength should be correlated in this fashion is intuitive if one thinks of these tests as assessing related portions of the same memory vector. There is unavoidable variability in how well individual items are globally encoded. Some items are well attended and encoded during study and hence yield higher strength on both recognition and source memory tests. Other items are less well attended and encoded during study and yield a lower memory strength on both tests. This would result in an item distribution for which the recognition and source memory dimensions are correlated. Moreover, the direction of this correlation would be such that items that have more strength on the recognition dimension also tend to have more strength in favor of their true source as well. Both of the old distributions in figure 3a-c exhibit such a correlation.

In both of the old distributions, the major axis of the ellipse (i.e., the longest possible line segment that can be enclosed within the ellipse) resides on the line that crosses that distribution's mean and the mean of the new distribution. That is, the direction in which the mean of the studied items has moved away from the mean of the unstudied items is the same direction that accounts for the most variability in that bivariate distribution (i.e., the major axis of that distribution's ellipse). Such a relationship, shown in Figure 4a, implies that the signal mean moves away from noise distribution in a direction that can be predicted by the covariance between the strengths on each test. This pattern would indicate that global memory factors, such as attention or general item difficulty, affect all areas of the memory vector at the same proportional rates. That is, each dimension in the vector has an equal probability of being encoded faithfully, regardless of whether that dimension includes information relevant to source discrimination or item recognition. However, this probability does vary from trial to trial, as evidenced by the presence of nonzero covariance.

#### *Test covariance and test selectivity*

If, however, there are factors that selectively enhance (or impair) memory on one test and not the other (i.e., factors that have differential effects on different regions of the memory vector), then the signal mean would move in a different direction than would be predicted by the covariance. Such a relationship, shown in Figure 4b, would suggest that there are selective factors that have disproportional effects on each test. For example, it is well known that word frequency affects item recognition (e.g., Benjamin, 2003; Glanzer & Adams, 1990) but the case for source recognition is less clear (Guttentag & Carroll, 1994; Marsh, Cook, & Hicks, 2006). To the extent that a factor can selectively



enhance (or impair) detection accuracy on one test and not the other, then the mean of the old distribution would move away from the noise distribution in a direction other than the one described by the major axes of its ellipse. The bivariate model presented here can directly evaluate this hypothesis about the nature of covariance in the signal distribution and compare it to predictions derived from a perspective assuming global effects of learning.

*Manipulations of memory and global accounts of learning and forgetting*

We now at long last return to the question of central interest here—how the path towards establishing memory taken during learning resembles and differs from the path taken during forgetting. The two experiments reported here employ a manipulation of learning (Experiment 1) and forgetting (Experiment 2) and investigate whether the movement of distributions reflects global or selective influences. These manipulations allow us to trace the effects of these manipulations with respect to the axis emanating from the noise distribution. Figure 4c shows the case where increases (or decreases) in memory strength continue along the direction predicted by the covariance. Figure 4d, on the other hand, shows the case in which the means move in a direction not predicted by the covariance, but nonetheless increase (or decrease) at the same proportional rate for all levels of the experimental manipulation. Finally, Figure 4e shows the case where the means are not predicted by the covariance and the proportional rate is different at each level of the experimental manipulation. In this model, a variable might affect item recognition more than source memory (for example) in one portion of the range and source memory more than item recognition in a another portion of the range.

In the experiments presented here, all of these models are fit to the data and the more flexible models (e.g., Figure 4e) are tested as to whether they provide a better fit than the less flexible models (e.g., Figure 4c and 4d). The goal of Experiment 1 is to fit a multivariate model to individual subject data, test its basic assumptions, evaluate the effects of a manipulation of learning on the model parameters, and compare it to the DPSD benchmark (Yonelinas, 1994, 1999) that is an alternative and still widely used model of source recognition performance.

### Experiment 1

This experiment sought to investigate how well the multivariate signal detection model is able to fit individual participant's recognition and source memory data and to compare those fits with the fits of the DPSD model. The fits were compared using the Akaike Information Criterion (AIC), which provides a goodness of fit measure that can be compared across non-nested models with different numbers of parameters. Subjects in this experiment engaged in both an item recognition (item memory) and source memory task for previously studied words.

This experiment also sought to investigate the relationship between recognition memory and source memory within the bivariate model. A study time manipulation was included to experimentally manipulate the strength of various items. To the extent that memory is strengthened globally by additional encoding, the proportional increase of the average item's strength on each test should be the same for items encoded better (i.e., items given more study time) and items encoded less well (i.e., items given less study time). This would be instantiated in the bivariate model by the means of the distributions of different levels of study time all moving away from the noise distributions along the

same vector. That is, the means should all fall on a single line, and that line should cross the mean of the noise distribution (see Figures 4c and 4d). Alternatively, if additional study time differentially affects retention of stimulus characteristics important for item memory and source memory, then the direction that the means move away from the noise distribution would change at higher levels of study time, and they would not fall on the same line (Figure 4e).

## Method

### *Participants*

Participants were 49 undergraduates enrolled in a psychology course at the University of Illinois at Urbana-Champaign.

### *Materials and Procedure*

Each participant studied 120 words presented one at a time on a computer screen. Half of the words were presented on the left side of the screen and the other half on the right side of the screen. One-third of the study words (20 in each location) were presented for 500 msec (weak encoding), one-third were presented for 3000 msec (medium encoding), and the remaining third were presented for 5000 msec (strong encoding). The order of presentation was pseudo-random with the constraint that no more than four words in a row could appear on the same side of the screen or for the same duration. Participants were instructed to try to remember each word and the location it was presented in for a later memory test.

The test consisted of all 120 studied words and 120 novel words. Each test item was presented by itself in the center of the screen. For each test item, participants first had to make a recognition judgment. The question “Was it studied?” appeared above the

test item and a 1 to 5 scale appeared below the test item, with 1 labeled “sure new” and 5 labeled “sure old.” Once the participant made a response on the recognition test, they were asked to make a source judgment on the same test item. The question “Where was the word on the screen?” then appeared above the test item and a 1 to 5 scale appeared below the test item with 1 labeled “sure left” and 5 labeled “sure right.” Participants were instructed prior to the test phase that if they think or know that a test item was not studied they should respond with the middle response on the scale for the latter question (i.e., the “3”). The presentation of test items was pseudo-randomized with the constraints that no more than four test items from the same studied location or duration or three novel test items appeared in a row.

#### *Model Comparisons*

Both the bivariate signal detection model and the DPSD model were fit for each participant simultaneously to both memory tasks using maximum likelihood estimation. A formal description of the models and of the fitting procedure can be found in the Appendix. Both models were fit with all parameters being free to vary (full models) and with parameters constrained to either be particular values or equal to other parameters (reduced models). The AIC was used to compare reduced and full models to determine if allowing a particular parameter to vary reliably improved the fit. The best fitting bivariate signal detection and DPSD models were then compared to each other using the AIC statistic. The models and their varying restrictions are summarized in Table 1, and described below.

*Multivariate models.* For the bivariate signal detection model, the full model allowed all the means and variance-covariance matrices of the 2 (location) X 3 (study

duration) old distributions to vary freely (the new distribution was constrained to be the z-distribution with mean (0,0) and variance-covariance equal to the identity matrix  $\mathbf{I}$ ). The various reduced models imposed constraints on either the means or the variance-covariance matrices or both.

There were four different levels of constraint for the means. The full (least reduced) model allowed the means to vary freely. The three reduced models all constrained the means of the three old distributions from the same location to lay on the same vector moving away from the new distribution mean (see Appendix). The most reduced model constrained the means of the same duration but different locations to be the same distance from the new distribution and for the vector for each location to form the same angle with the item recognition axis but in opposite directions. That is, the means from the same duration but opposite locations had the same value of recognition strength but additive inverse source strength. The next most reduced model constrained the vectors but allowed each distribution's distance from the new distribution to vary freely. Finally, the last reduced model in terms of the means constrained the means to fall on a vector for each location, but allowed the direction of those vectors to vary freely.

These four levels of mean constraints were fully crossed with two levels of covariance restrictions, yielding eight total models. The full model allowed the variance-covariance matrices of the signal distributions to vary freely. The reduced model constrained the variance-covariance matrix such that the direction of the major axes of the ellipse was the same as the line going from the mean of the new distribution to that distribution's mean. That is, the direction that accounts for the greatest variance in the

distribution was constrained to be the same as the direction in which the mean of that distribution has “moved away” from the new distribution.

*Dual-process models.* There were three different sets of  $\underline{R}$  and  $\underline{d}$  parameters—one set for each location and study duration. The various reduced models constrained those parameters within each set, but never constrained the parameters to be equal across items of different durations. The full model had a separate  $\underline{R}$  parameter and a separate  $\underline{d}$  parameter for each item type (left vs. right) during each test type (recognition vs. source). There were three different reduced models for both the  $\underline{R}$  and  $\underline{d}$  parameters. For each parameter type, one constrained model forced that parameter to be the same for both test types but allowed it to vary independently for each location. Another reduced model constrained that parameter to be the same for each location but allowed it to differ for the two test types. Finally, the strictest model constrained both parameters to be the same for all item types and test types within a study duration. This yielded four different levels of constraint for both the  $\underline{R}$  and the  $\underline{d}$  parameters. Fully crossing all levels of  $\underline{R}$  constraint and  $\underline{d}$  constraint yielded 16 different models that were fit. These relaxed models afford the DPSD more flexibility than theorists typically allow it—e.g., recollection differing for items from a different source and depending on the test format—but we wanted to give the model every opportunity to fit the data. The most flexible model had a different  $\underline{R}$  and a different  $\underline{d}$  parameter for every old item type on each test. Yonelinas (1999) specifically argues that  $\underline{d}$  should be the same for all studied item types when the items are all equally memorable. Likewise, there is no theoretical reason why  $\underline{R}$  would differ for the exact same items on different tests, or differ for items that were studied for equivalent

time but in different locations. Nonetheless, we allowed these parameters to vary freely to test possible variants of dual-process models.

## Results

Prior to fitting the model to both tasks simultaneously, we examined the properties of the univariate fits of the models so as to confirm the effectiveness of the study time manipulation and to ensure that basic effects from the prior literature were replicated in our task and with our model-fitting algorithms.

### *Univariate analysis of performance*

The mean parameter estimates for the univariate fits of both the UVSD and DPSD model can be found in Table 2. For the UVSD model,  $\underline{d}_a$  reliably increased from the weak condition to the medium condition for both the recognition task and source memory task ( $t(48) = 3.29$ ,  $t(48) = 2.63$ , respectively). This indicates that the study time manipulation did have its desired effect. However, there was no reliable increase between the medium and strong conditions ( $t(48) = .97$ ,  $t(48) = .05$ , for recognition and source memory, respectively). Figure 5 shows the ROC plots and the group fit<sup>3</sup> UVSD curves for both the recognition and source memory tasks.

The DPSD model also found a reliable increase between the weak and medium word types but not the medium and strong word types, though this pattern only occurred for the  $\underline{R}$  parameter in the source memory task ( $t(48) = 3.62$ ,  $t(48) = .81$ , for the weak vs. medium and medium vs. strong comparisons, respectively). The increase in the DPSD parameters for the recognition test was not statistically reliable between the weak and medium conditions ( $t(48) = 1.54$  &  $t(48) = .33$ , for the  $\underline{d}$  and  $\underline{R}$  parameters, respectively)

---

<sup>3</sup> All parameter estimates and statistical tests done on those parameter estimates were conducted on an individual participant basis. The group-fit ROC curves are plotted for clarity of presentation.

or between the medium and strong conditions ( $t(48) = 1.05$  &  $t(48) = 1.06$ , for the  $\underline{d}$  and  $\underline{R}$  parameters, respectively). However, there was a reliable increase between the weak condition and the strong condition for the  $\underline{d}$  parameter ( $t(48) = 2.92$ ) and the increase between those conditions for the  $\underline{R}$  parameter was marginally reliable ( $t(48) = 1.76$ ,  $p=.08$ ), indicating the DPSD fits also support the notion that the memory strength manipulation did have the desired effect. The ROC plots and the group-fit DPSD curves for both the recognition and source memory tasks are shown in Figure 6.

### *Bivariate Model Fits*

The mean AIC values across participants for the various versions of the bivariate signal detection model that were fit are presented in Table 3. The best fitting model was the one with the fewest constraints—both the means and the covariances were allowed to vary freely. Figure 7 shows the fit of the best fitting model using the median parameter estimates across participants.

The main focus of this experiment was to evaluate whether what the fitted parameters from the multivariate model indicate about the relationship between study time and learning. One possibility was that the average item's strength on each test would increase at the same proportional rate as memory improves. If this were the case, then all the means from the same source would fall on the same line that crosses the noise distribution. The best fitting model however, allowed the means to vary freely from this line. The models that constrained the means to be on the same line did not fit the data as well. This suggests that the effects of encoding are not equivalent across all areas of the memory trace. Rather, it appears that recognition strength benefits most from the initial encoding, and source strength benefits most from the increases in study time (Figure 7).



This result is consistent with the “one-shot” hypothesis of Malmberg and Shiffrin (2005) and will be discussed further in the General Discussion. In addition, allowing the variance-covariance matrix to vary freely fit better than when that matrix was constrained. This suggests that there are factors that affect the strength on one task but not the other.

#### *DPSD model fit simultaneously*

The mean AIC values for each of the DPSD models fit are shown in Table 4. The best fitting model was the model with the same  $\underline{R}$  parameter for each study location but a different  $\underline{R}$  parameter for each test. This model implies that words studied in each context are recollected at the same rate, on average, but that the probability of recollection differs depending on the type of test. This model did not do as good a job fitting the data as the bivariate model, indicating that recognition and source memory are best fit by a continuous strength model, rather than a mixture of a strength and threshold model. This is inconsistent with the theorizing of theories that suggest that familiarity (or recollection) should be the same for equally studied items and for the same item on two different tests.

#### Discussion

This experiment sought to assess the effect of increasing study time on the memory representation underlying performance on recognition and source memory. The DPSD model fit simultaneously to both tasks failed to fit the data as well as the bivariate model. This supports the notion that recognition and source memory are best fit by a continuous strength process.

The results of the bivariate modeling do not support the hypothesis that the signal underlying recognition and the signal underlying source memory increase proportionately and instead indicate that recognition and source strength benefit differentially at different levels of encoding. If one were to fit the best fitting line to the three means of items from the same study context (Figure 7), that line would intercept the item recognition axes well above the mean of the noise distribution. This result suggests that the source strength benefits most from increases in study time, and that recognition strength benefited most from the initial encoding. One possibility is that participants are using the extra study time to better associate the item with its location. This explanation is consistent with the notion that encoding does not affect all areas of the memory trace equally, but rather, that participants deliberately attend to some aspects of the stimulus over others.

## Experiment 2

Experiment 1 found that as the strength of an item's encoding is increased, the memory strength on each test also increases, though not at the same proportional rate. This finding is inconsistent with the idea that increases in study time globally strengthen the entire memory trace. If encoding strengthened all areas of the memory trace globally, then we would expect the means of studied item distributions to move away from the mean of the noise distribution along a straight line. This was not the case in Experiment 1 when we used a study time manipulation to differentially strengthen the items.

Experiment 2 sought to manipulate item strength using a forgetting manipulation. This allowed us to evaluate whether the effects of forgetting mirror the effects of study time: If so, source performance should decrease more rapidly at low levels of forgetting, and item performance should show a rapid decline later. However, to the extent that

forgetting, unlike study time, affects a memory trace globally, we would expect the means of studied items that are more likely to have been forgotten to move towards the mean of the noise distribution along a straight line.

## Method

### *Participants*

Participants were 66 undergraduates enrolled in a psychology course at the University of Illinois, Urbana-Champaign.

### *Materials and Procedure*

Materials were the same as Experiment 1. The procedure was also the same as Experiment 1, except that all items were studied for two seconds and there were three separate test phases. Each test queried a random but mutually exclusive third of the study words (20 from each location). The “strong” words were tested immediately after the study phase. For the “medium” words participants were asked to complete as many 2-digit addition and subtraction problems as they could in 20 seconds immediately prior to the second study phase. For the “weak” words participants were asked to solve as many of those math problems as they could in 180 seconds immediately prior to the third and final test phase. The math problems consisted of two randomly generated integers each ranging from 10 to 99 and a randomly selected “+” or “-“ sign presented on a computer. For the subtraction problems, the two integers were sorted such that the correct answer was never negative.

### *Model Simulations*

The same models were fit as in Experiment 1. The fitting algorithm failed to achieve good fits for two of the participants. The best fits for those participants were

more than 6 standard deviations worse than the other 64 participants, so the bivariate fits for those two participants were not included in the results below.

## Results

### *Univariate analysis of performance*

The mean parameter estimates for the univariate fits of both the UVSD and DPSD model can be found in Table 5. For the UVSD model,  $\underline{d}_a$  reliably decreased from the strong condition to the weak condition for both the recognition task and source memory task ( $t(65) = 2.10$ ,  $t(65) = 2.36$ , respectively) indicating that the memory strength of the items decreased on both tests as the items were forgotten. Figure 8 shows the ROC plots and the group fit UVSD curves for both the recognition and source memory tasks.

The DPSD model also detected a reliable decrease from the strong to the weak condition, but only in the  $\underline{R}$  parameter during the recognition task ( $t(65) = 3.08$ ). Neither the decrease of the  $\underline{d}$  parameter during the recognition task nor of the  $\underline{R}$  parameter during the source memory task were statistically reliable ( $t(65) = .20$ ,  $t(65) = .86$ , respectively). This pattern of results appears to indicate that recollection is more susceptible to forgetting than familiarity and supports the notion that the experimental manipulation led to forgetting. The ROC plots and the group fit DPSD curves for both the recognition and source memory tasks are shown in Figure 9.

### *Bivariate Model Fits*

The mean AIC values across participants for the various versions of the bivariate signal detection model that were fit are presented in Table 6. Unlike the results of Experiment 1, the model that allowed the means to vary freely did not provide a better fit than the model that constrained the means to fall on the same line. This result is

consistent with the notion that forgetting occurs globally across the memory trace. That is, memory strength decreased at a proportional rate across the two tasks. As items were weakened through forgetting, the mean of the corresponding distribution moved along a straight line towards the mean of noise distribution. Figure 10 shows the median fit of the best fitting model.

As in the previous experiment, the models that allowed the variance-covariance matrix to vary freely fit better than when that matrix was constrained. This again suggests that there are factors that affect the strength on one task but not the other. This is in contrast to the main finding of this experiment, that forgetting acts globally across the memory trace. Forgetting affects the strength on both tests at a proportional rate, but there are still other factors (presumably the same as those elicited across heterogeneous items and encoding opportunities in experiment 1) that affect the strength on one test differently than the strength on another test.

#### *DPSD model fit simultaneously*

The mean AIC values for each of the DPSD models fit are shown in Table 7. The best fitting model was the same as in Experiment 1; the model with the same  $\underline{R}$  parameter for each study location but a different  $\underline{R}$  parameter for each test. This model implies that words studied in each context are recollected at the same rate, on average, but that the probability of recollection rate differs depending on the type of test. The best fitting DPSD model once again failed to fit the data as well as the best fitting bivariate model, though overall the DPSD model fit the data better in this experiment than in the previous experiment.

## Discussion

Experiment 1 found that the effect of strengthening items, through the use of a study time manipulation, affected the strength of those items on the two memory tests at different proportional rates. This experiment investigated whether the effects of weakening items, through the use of a forgetting manipulation, affected the strength of those items at the same proportional rate for the two memory tests. The results of the bivariate modeling support this hypothesis and suggest that forgetting occurs globally across the memory trace. The model that constrained the decreases in the source mean to be a constant proportion of the decreases in the recognition mean (i.e., for the means to fall on a straight line that crosses the origin) provided a better fit than the model that allowed the means to vary freely. This is consistent with the view that the amount of forgetting in areas assessed by the source test is a constant proportion of the amount of forgetting that occurs in areas assessed by the recognition test.

Once again, the bivariate model did a better job accounting for the data than did the DPSD model. This result, consistent across the two experiments and in other recent work (Hautus et al., 2008), supports the viability of a continuous strength model as an explanation for recognition and source memory performance.

### General Discussion

The experiments presented here explored similarities and differences between learning and forgetting by using a continuous strength model to simultaneously account for performance on tests of recognition memory and tests of source memory. The model developed here is grounded in the source monitoring framework that suggests that an item's memory strength depends upon what aspects of memory are being queried. Items that are learned equally well may yield equal memory strengths on one memory test, but

entirely different memory strengths on a different test. The bivariate model fit in the experiments presented here found that items that were studied for equal durations but in different locations had approximately equal memory strength on the recognition task, but nonetheless elicited well above-chance source memory discrimination performance. In the multivariate formulation, this can occur because the areas of memory that are assessed for the recognition test contain similar information for both types of studied words, whereas the areas of memory assessed for the source test contain very different information for the two types of studied words.

The bivariate model's ability to fit both tasks simultaneously, and to do so better than the DPSD model, goes against the notion that source memory performance depends critically upon recollection and that recollection differs qualitatively from strength in that it is a threshold process (cf. Wixted, 2007). By allowing an item's strength to depend upon what aspect of that memory trace is being tested, source memory performance can be well described by a familiarity or strength process. This basic assumption of the bivariate model presented here follows directly from the source monitoring framework. According to that framework, the type of evidence that is assessed in memory depends upon what is relevant to the test question. Items that are learned equally well can yield drastically different levels of evidence if the test is one that queries aspects of memory where the items differ, such as a source memory test.

#### *Properties of Bivariate fits*

The bivariate model enabled us to evaluate the rate at which the average item's strength changed proportionately on the two tests as memory for the stimuli was strengthened (Experiment 1) or weakened (Experiment 2). The results of Experiment 1

revealed that the rate at which item strengths increase on the two tests does not remain constant as the study time is increased. Initially, strength increases most along the recognition dimension. Then, as participants are given more time to study the items, the rate at which the strength increases across the two tests changes in favor of the source memory dimension. Moreover, this latter direction of movement appears to be roughly in the same direction as predicted by the major axis of the ellipse (Figure 7), indicating that the same factors that underlie item-to-item variability in encoding efficacy promote enhancements in source and item memory, but only after an initial period of study time for which item recognition increases disproportionately.

This pattern may be consistent with the one-shot hypothesis of Malmberg and Shiffrin (2005). According to that hypothesis, all the information that will be stored about the context of a study episode is stored in “one shot” at the initial presentation and all subsequent study time is spent encoding the “content” of the study item. While this hypothesis appears at first glance to be the opposite of what we found, it is worth noting that Malmberg and Shiffrin (2005) conceptualize context as the situational aspects of a study episode that are usually outside the focus of attention during deliberate study. Since participants were told to remember the location of the word while they studied that word, it is more appropriate to consider location in our experiments as part of the “content” of the item. Malmberg and Shiffrin (2005) model context in REM (Shiffrin & Steyvers, 1997) as a section of the memory trace that is unique to each particular encoding episode (but that does not get better encoded as study time increases). Such a model can account for why recognition strength increases initially more so than predicted by the covariance, but later with more study time, strength on both tests appear to move



in the direction predicted by the covariance. The unique context information for each word aids in making recognition judgments for those words, but is of little use in deciding where that word was on the screen. This initially boosts performance disproportionately in favor the recognition test. Since that context information does not benefit from more study time however, later increases in strength from additional study tend to move in the direction predicted by the covariance.

The results of Experiment 2 are quite different than Experiment 1. When item strength was weakened through forgetting, strength on both the recognition and source memory tests decreased by the same proportional amount. That is, the strength of the average item moved in a straight line towards the mean of the noise distribution as items are forgotten. This suggests that forgetting occurs globally across the memory trace and does not differentially affect some areas more than others. It also serves to allay concerns that the results of the previous experiment were due to the more flexible model overfitting the data. This experiment shows that the model where the means fall on a straight line is flexible enough to provide the best fit to the data.

Overall, the bivariate model was able to account for performance on both recognition memory and source memory tests, outperformed the DPSD model, and allowed us to explore the relation of recognition strength to source strength. The results were consistent with the one-shot hypothesis that initially there are factors the preferentially aid recognition memory and later increases in study time affect both recognition and source memory strength, but that forgetting is more egalitarian in its effects on the memory trace.

## References

- Ashby, F. G. (1992). *Multidimensional models of perception and cognition*. Hillsdale, NJ: Erlbaum.
- Atkinson, R. C., & Juola, J. F. (1973). Factors influencing speed and accuracy of word recognition. In S. Kornblum (Ed.), *Attention and performance IV*. New York: Academic Press.
- Atkinson, R. C., & Juola, J. F. (1974). Search and decision processes in recognition memory. In D. H. Krantz, R. C. Atkinson, R. D. Luce, & P. Suppes (Eds.), *Contemporary developments in mathematical psychology (Vol. 1): Learning, memory, and thinking*. San Francisco: Freeman.
- Banks, W. P. (2000). Recognition and source memory as multivariate decision processes. *Psychological Science, 11*, 267–273
- Benjamin, A. S. (2003). Predicting and postdicting the effects of word frequency on memory. *Memory & Cognition, 31*, 297-305.
- Benjamin, A.S., Diaz, M., & Wee, S. (2009). Signal detection with criterion noise: Applications to recognition memory. *Psychological Review, 116*, 84-115.
- DeCarlo, L. T. (2002). Signal detection theory with finite mixture distributions: Theoretical developments with applications to recognition memory. *Psychological Review, 109*, 710-721.
- DeCarlo, L. T. (2003a). An application of signal detection theory with finite mixture distributions to source discrimination. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*, 767-778.

- DeCarlo, L. T. (2003b). Source monitoring and multivariate signal detection theory, with a model for selection. *Journal of Mathematical Psychology*, *47*, 292–303.
- Egan, J. P. (1958). *Recognition memory and the operating characteristic* (Technical Note No. AFCRC-TN-58–51). Bloomington: Indian University, Hearing and Communication Laboratory.
- Glanzer, M. & Adams (1990). The mirror effect in recognition memory. *Memory & Cognition*, *13*, 8-20.
- Glanzer, M., Hilford, A., & Kim, K. (2004). Six Regularities of Source Recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(6), 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.
- Green, D. M., & Swets J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.
- Haskins, A. L., Yonelinas, A. P., Quamme, J. R., & Ranganath, C. (2008). Perirhinal cortex supports encoding and familiarity-based recognition of novel associations. *Neuron*, *59*, 554–560.
- Hautus, M., Macmillan, N. A., & Rotello, C. M. (2008). Toward a complete decision model of item and source recognition. *Psychonomic Bulletin & Review*, *15*, 889-905.
- Healy, M. R., Light, L. L., & Chung, C. (2005). Dual-process models of associative recognition in young and older adults: Evidence from receiver operating

- characteristics. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31, 768-788.
- Heathcote, A. (2003). Item recognition memory and the ROC. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29, 1210-1230.
- Hilford, A., Glanzer, M., Kim, K., & DeCarlo, L. T. (2002). Theories of source memory: ROC analysis. *Journal of Experimental Psychology: General*, 131, 494–570.
- Hintzman, D. L. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review*, 93, 411-428.
- Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language*, 30, 513–541.
- Jacoby, L. L., Shimizu, Y., Daniels, K. A. & Rhodes, M. G. (2005). Modes of cognitive control in recognition and source memory: Depth of retrieval. *Psychonomic Bulletin and Review*, 12, 852-857.
- Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin*, 114, 3–28.
- Kelley, R., & Wixted, J. T. (2001). On the nature of associative information in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27, 701–722.
- Macho, S. (2002). Cognitive modeling with spreadsheets. *Behavior Research Methods, Instruments & Computers*, 34, 19-36.
- Macho, S. (2004). Modeling associative recognition: A comparison of two-high-threshold, two-high-threshold signal detection, and mixture distribution models.

*Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30, 83-97

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

Malmberg, K. J. & Shiffrin, R. M. (2005). The “one-shot” hypothesis for context storage. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31, 322-336.

Mandler, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological Review*, 87(3), 252-271. doi:10.1037/0033-295X.87.3.252.

Marsh, R. L., Cook, G. I., & Hicks, J. L. (2006). The effect of context variability on source memory. *Memory & Cognition*, 34, 1578-1586.

Marsh, R. L. & Hicks, J. L. (1998). Test formats change source-monitoring decision processes. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 24, 1137-1151.

Matzen, L.E., and Benjamin, A.S. (2009). Remembering Words Not Presented in Sentences: How Study Context Changes Patterns of False Memories. *Memory & Cognition*, 37(1), 52-64.

Mickes, L., Johnson, E. M., & Wixted, J. T. (2010). Continuous recollection versus unitized familiarity in associative recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36(4), 843-863.  
doi:10.1037/a0019755

Murdock, B. B., Jr. (1982). A theory for the storage and retrieval of item and associative information. *Psychological Review*, 89, 609–626.

- Onyper, S. V., Zhang, Y. and Howard, M. W. (2010). Some-or-none recollection: Evidence from item and source memory. *Journal of Experimental Psychology: General*, *139*, 341-364.
- Parks, T. E. (1966). Signal-detectability theory of recognition performance. *Psychological Review*, *73*, 44–58.
- Qin, J., Raye, C. L., Johnson, M. K., & Mitchell, K. J. (2001). Source ROCs are (typically) curvilinear: Comment on Yonelinas (1999). *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*, 1110–1115.
- Ranganath, C., & Paller, K. A. (2000). Neural correlates of memory retrieval and evaluation. *Cognitive Brain Research*, *9*, 209-222.
- Ratcliff, R., McKoon, G., & Tindall, M. (1994). Empirical generality of data from recognition memory receiver-operating characteristic functions and implications for the global memory models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 763-785.
- Ratcliff, R., Sheu, C. F., & Gronlund, S. D. (1992). Testing global memory models using ROC curves. *Psychological Review*, *99*, 518-535.
- Rugg, M. D., Allan, K., & Birch, C. S. (2000). Electrophysiological evidence for the modulation of retrieval orientation by depth of study processing. *Journal of Cognitive Neuroscience*, *12*, 664-678.
- Senkfor, A. J., & Van Petten, C. (1998). Who said what? An event-related potential investigation of source and item memory. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *24*, 1005-1025.

- Sherman, S. J., Atri, A., Hasselmo, M. E., Stern, C. E., & Howard, M. W. (2003). Scopolamine impairs human recognition memory: Data and modeling. *Behavioral Neuroscience, 117*, 526-539.
- Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: REM: Retrieving effectively from memory. *Psychonomic Bulletin & Review, 4*, 145–166.
- Slotnick, S. D., & Dodson, C. S. (2005). Support for a continuous (single-process) model of recognition memory and source memory. *Memory & Cognition, 33*, 151-170.
- Slotnick, S. D., Klein, S. A., Dodson, C. S., & Shimamura, A. P. (2000). An analysis of signal detection and threshold models of source memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 1499–1517.
- Swets, J. A. (1963). Information retrieval systems. *Science, 141*, 245-247.
- Van Petten, C., Senkfor, A. J., & Newberg, W. M. (2000). Memory for drawings in locations: Spatial source memory and event-related potentials. *Psychophysiology, 37*, 551-564.
- Wickens, T. D. (1992). Maximum-likelihood estimation of a multivariate Gaussian rating model with excluded data. *Journal of Mathematical Psychology, 36*, 213–234.
- Wixted, J. T. (2007). Dual-process theory and signal-detection theory of recognition memory. *Psychological Review, 114*, 152–176.
- Yonelinas, A. P. (1994). Receiver-operating characteristics in recognition memory: Evidence for a dual-process model. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 1341–1354.

- Yonelinas, A. P. (1997). Recognition memory ROCs for item and associative information: The contribution of recollection and familiarity. *Memory and Cognition*, 25, 747–763.
- Yonelinas, A. P. (1999). The contribution of recollection and familiarity to recognition and source-memory judgments: A formal dual-process and an analysis of receiver operating characteristics. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25, 1415–1434.
- Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language*, 46, 441–517.
- Yonelinas, A. P., & Parks, C. M. (2007) Receiver operating characteristics (ROCs) in recognition memory: A review. *Psychological Bulletin*, 133, 800-832.



#### Author note

Michael Diaz is now at Washington University in St. Louis. Funding was provided by an NSF graduate fellowship to Michael Diaz and grant R01 AG026263 to Aaron Benjamin.

We thank the members of the Human Memory and Cognition Laboratory at the University of Illinois for their helpful feedback on this work.

Table 1

*Versions of the bivariate signal detection and DPSD models that were fit*

Bivariate Model	Constraint on Means	Constraint on Covariance
Full	None	None
SameLineFullCovariance	Means on Line	None
SameLine&AngleFullCovariance	Means on Line, same angle	None
SameMeansFullCovariance	Same means, inverse source	None
FullSameVector	None	Cov constrained
SameLineSameVector	Means on Line	Cov constrained
SameLine&AngleSameVector	Means on Line, same angle	Cov constrained
SameMeanSameVector	Same means, inverse source	Cov constrained

DPSD Model	Constraint on R	Constraint on $d'$
Full	None	None
Same $D_{item}$ FreeR	None	$d$ same for both locations
Same $D_{test}$ FreeR	None	$d$ same for both tests
SameDFreeR	None	$d$ same for both locations and tests
FreeDSame $R_{item}$	R same for both locations	None
Same $D_{item}$ & $R_{item}$	R same for both locations	$d$ same for both locations
Same $D_{test}$ & $R_{item}$	R same for both locations	$d$ same for both tests
SameD& $R_{item}$	R same for both locations	$d$ same for both locations and tests
FreeDSame $R_{test}$	R same for both tests	None
Same $D_{item}$ & $R_{test}$	R same for both tests	$d$ same for both locations
Same $D_{test}$ & $R_{test}$	R same for both tests	$d$ same for both tests
SameD& $R_{test}$	R same for both test	$d$ same for both locations and tests
FreeDSameR	R same for both locations and tests	None
Same $D_{item}$ &R	R same for both locations and tests	$d$ same for both locations
Same $D_{test}$ &R	R same for both locations and tests	$d$ same for both tests
SameD&R	R same for both locations and tests	$d$ same for both locations and tests

Table 2

*Mean parameter estimates for each level of study time for Experiment 1*

Test	<u>UVSD</u>			<u>DPSD</u>	
	mu	sig	d <sub>a</sub>	d	R
<u>Old vs. new</u>					
Weak	1.25	1.62	.87	.48	.24
Medium	1.39	1.59	1.02	.60	.25
Strong	1.38	1.49	1.07	.67	.28
<u>Source A vs. B</u>					
Weak	.76	1.06	.72	0 <sup>a</sup>	.17
Medium	.99	1.16	.93	0 <sup>a</sup>	.24
Strong	.98	1.05	.93	0 <sup>a</sup>	.26

Note. These are the means of the parameters of the univariate models estimated individually for each participant. “Old vs. new” represents the comparison of old to novel items during the recognition task. “Source A vs. B” represents the comparison of the two different sources during the source memory task. a: d between the two source distributions is constrained to be 0 (as in Yonelinas [1999]).

Table 3

*Mean AIC values for various model fits of the Bivariate Signal Detection Model in Experiment 1*

---

Constrained Cov	<u>Constrained Means</u>			
	Full	Same Line	Same Line and Angle	Same Means
Full	<b>409</b>	411	701	718
Same Vector	957	964	813	828

---

Note. Lower values indicate better fit. The best fitting model is indicated in bold.

Table 4

*Mean AIC values for various model fits of the DPSD Model in Experiment 1*

Constrained R	<u>Constrained d'</u>			
	Full	Same d' for items	Same d' for tests	Single d'
Full	447	451	452	456
Same <u>R</u> for items	<b>442</b>	446	446	452
Same <u>R</u> for tests	459	463	465	470
Single <u>R</u> parameter	458	463	466	486

Note. Lower values indicate better fit. The best fitting model is indicated in bold.

Table 5

*Mean parameter estimates for each level of study time for Experiment 2*

Test	<u>UVSD</u>			<u>DPSD</u>	
	mu	sig	d <sub>a</sub>	d	R
<u>Old vs. new</u>					
Weak	1.84	2.45	1.01	.58	.28
Medium	1.86	2.37	1.07	.55	.33
Strong	1.98	2.14	1.15	.56	.38
<u>Source A vs. B</u>					
Weak	.81	1.06	.72	0 <sup>a</sup>	.13
Medium	.82	1.05	.80	0 <sup>a</sup>	.13
Strong	.99	1.19	.88	0 <sup>a</sup>	.16

Note. These are the means of the parameters of the univariate models estimated individually for each participant. “Old vs. new” represents the comparison of old to novel items during the recognition task. “Source A vs. B” represents the comparison of the two different sources during the source memory task.  $a:d$  between the two source distributions is constrained to be 0 (as in Yonelinas [1999]).

Table 6

*Mean AIC values for various model fits of the Bivariate Signal Detection Model in Experiment 2*

---

	<u>Constrained Means</u>			
Constrained Cov	Full	Same Line	Same Line and Angle	Same Means
Full	383	<b>379</b>	671	679
Same Vector	484	485	1427	2627

---

Note. Lower values indicate better fit. The best fitting model is indicated in bold.

Table 7

*Mean AIC values for various model fits of the DPSD Model in Experiment 2*

Constrained R	<u>Constrained d'</u>			
	Full	Same d' for items	Same d' for tests	Single d'
Full	387	394	393	401
Same <u>R</u> for items	<b>384</b>	391	390	399
Same <u>R</u> for tests	393	400	400	413
Single <u>R</u> parameter	395	402	408	424

Note. Lower values indicate better fit. The best fitting model is indicated in bold.



## Figure Captions

Figure 1

- a) Equal-variance model of recognition.

Figure 2

- a) ROC curves predicted by equal-variance (solid) and unequal-variance (dashed) strength models.
- b) z-transforms of ROC curves predicted by strength models
- c) ROC curve predicted by DPSD model
- d) z-transform of ROC curve predicted by DPSD model

Figure 3

- a) Bivariate signal detection model
- b) Horizontal plane passing through bivariate model
- c) Iso-likelihood contours of bivariate model

Figure 4

Potential patterns of means and covariance fits.

Figure 5

UVSD fits to the recognition and source memory group data

Figure 6

DPSD fits to the recognition and source memory group data

Figure 7

Iso-likelihood contour plot of the best fitting bivariate model in Experiment 1. Parameters used were the median parameter values across participants

Figure 8

UVSD fits to the recognition and source memory group data

Figure 9

DPSD fits to the recognition and source memory group data

Figure 10

Iso-likelihood contour plot of the best fitting bivariate model in Experiment 2. Parameters used were the median parameter values across participants

Figure 1

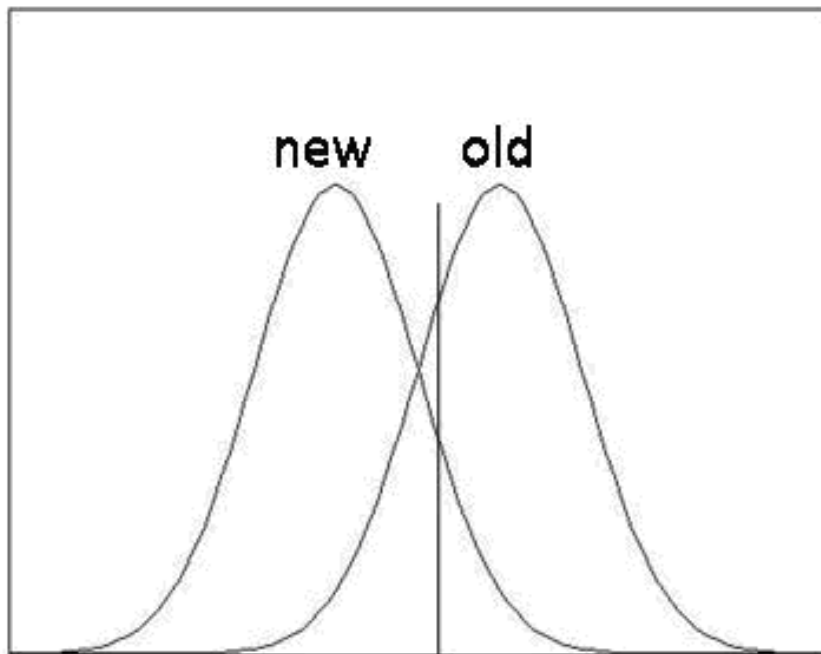


Figure 2a

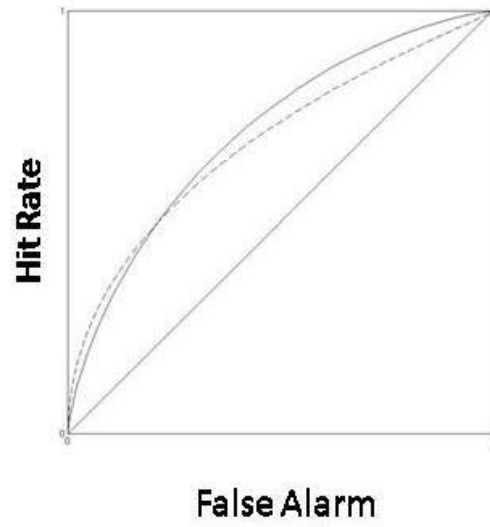


Figure 2b

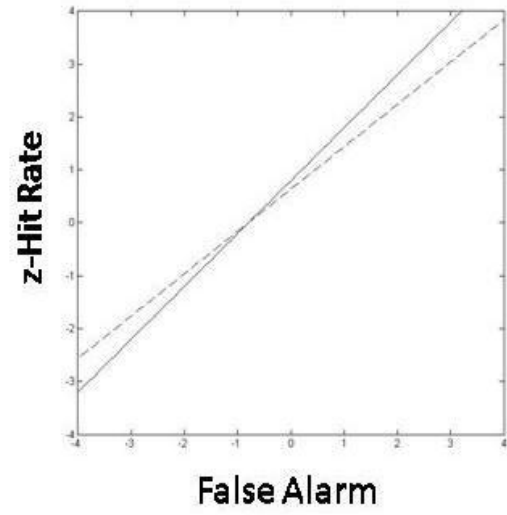


Figure 2c

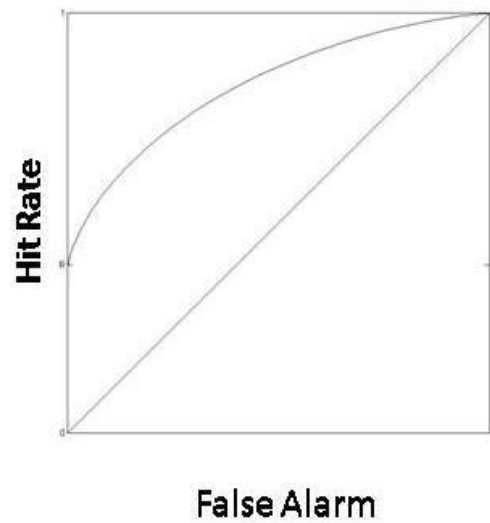


Figure 2d

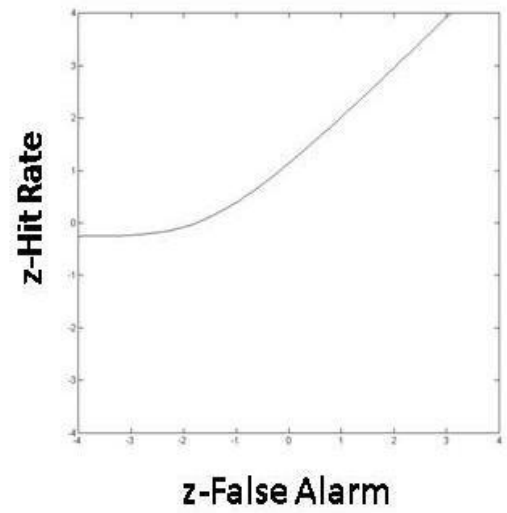


Figure 3

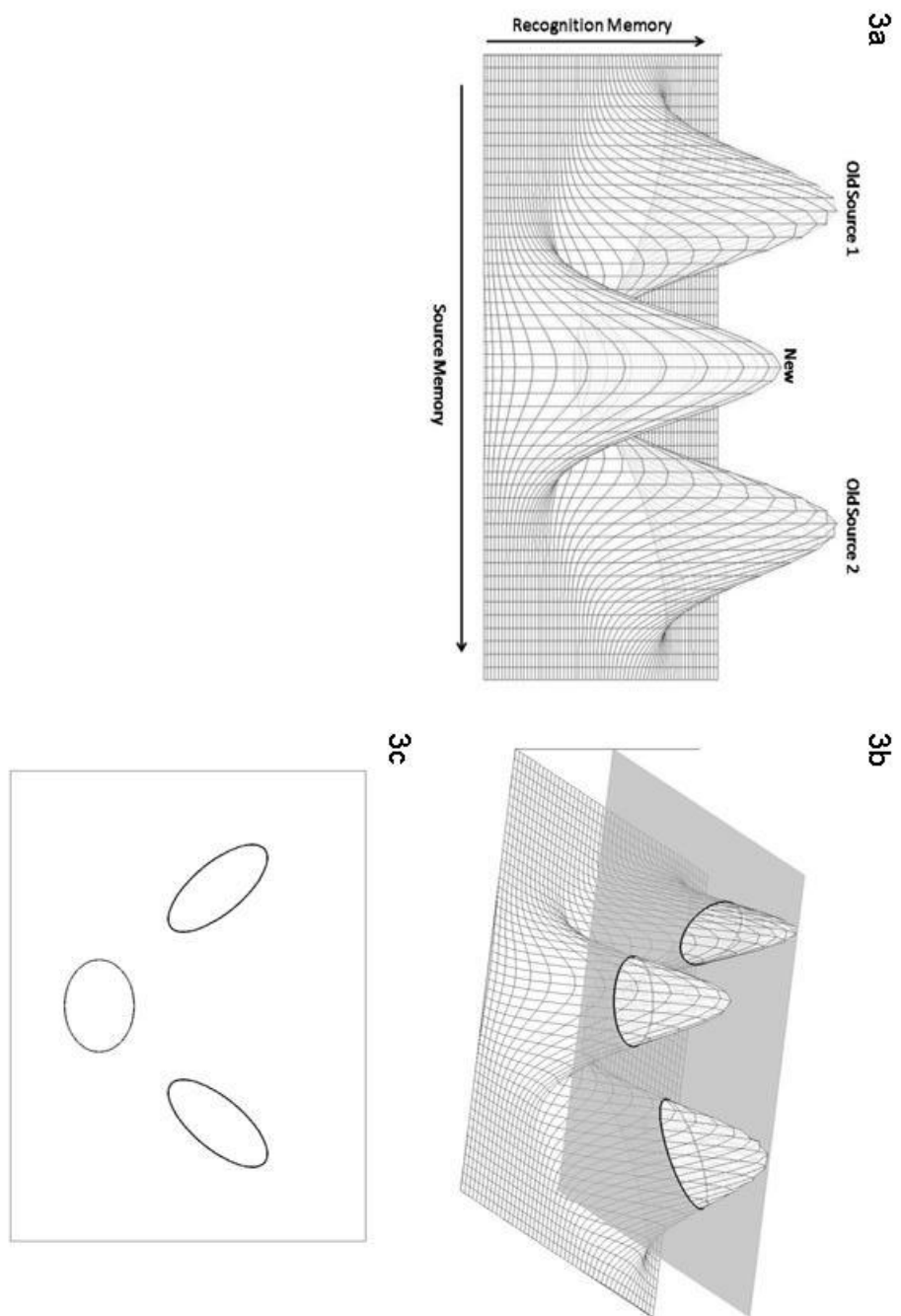


Figure 4

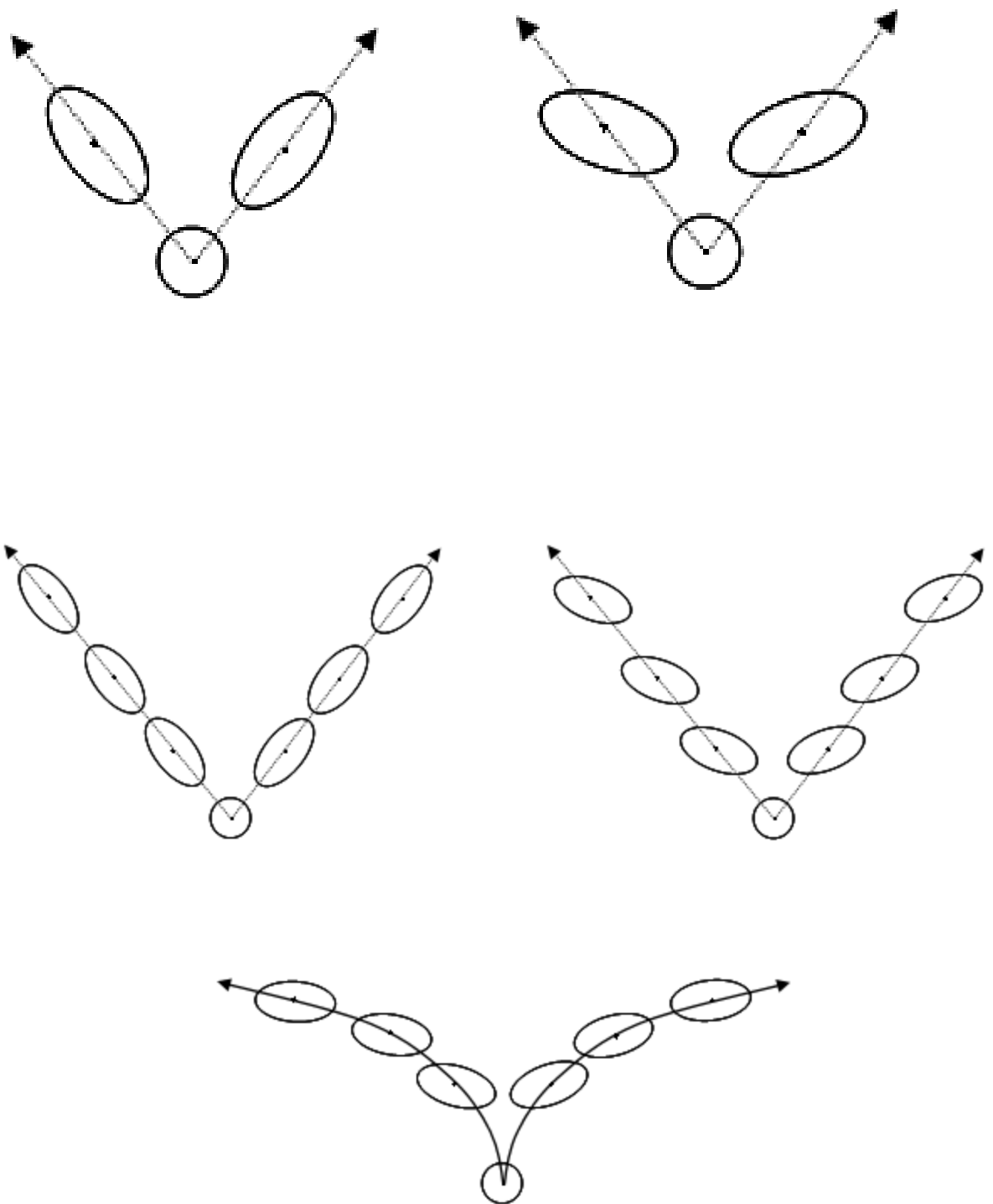
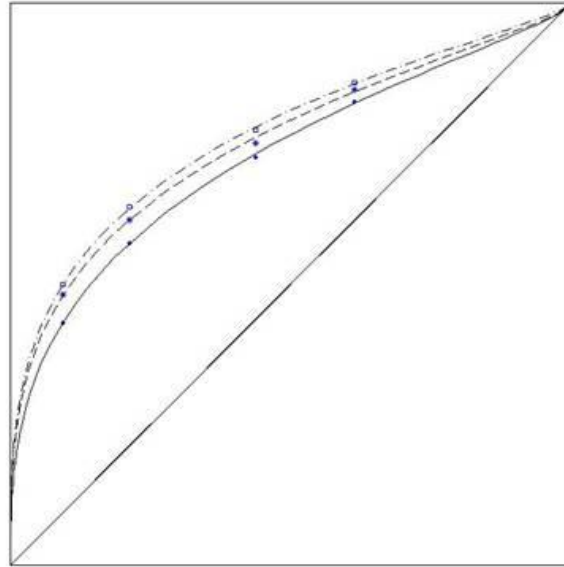


Figure 5

## Recognition



## Source Memory

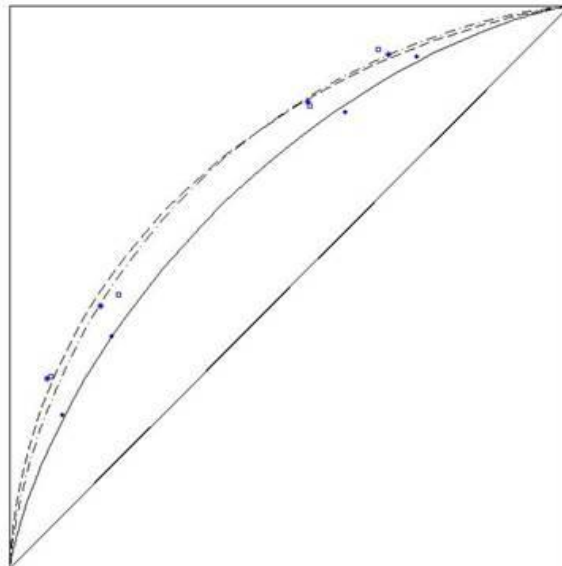
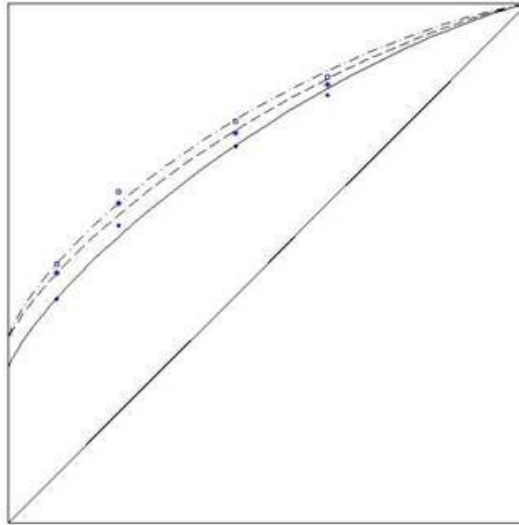


Figure 6

## Recognition



## Source Memory

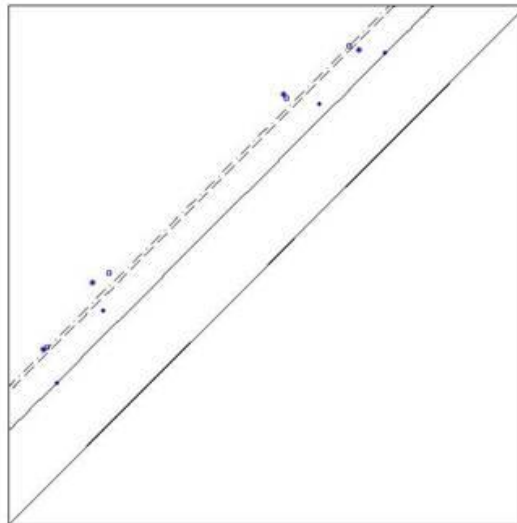


Figure 7

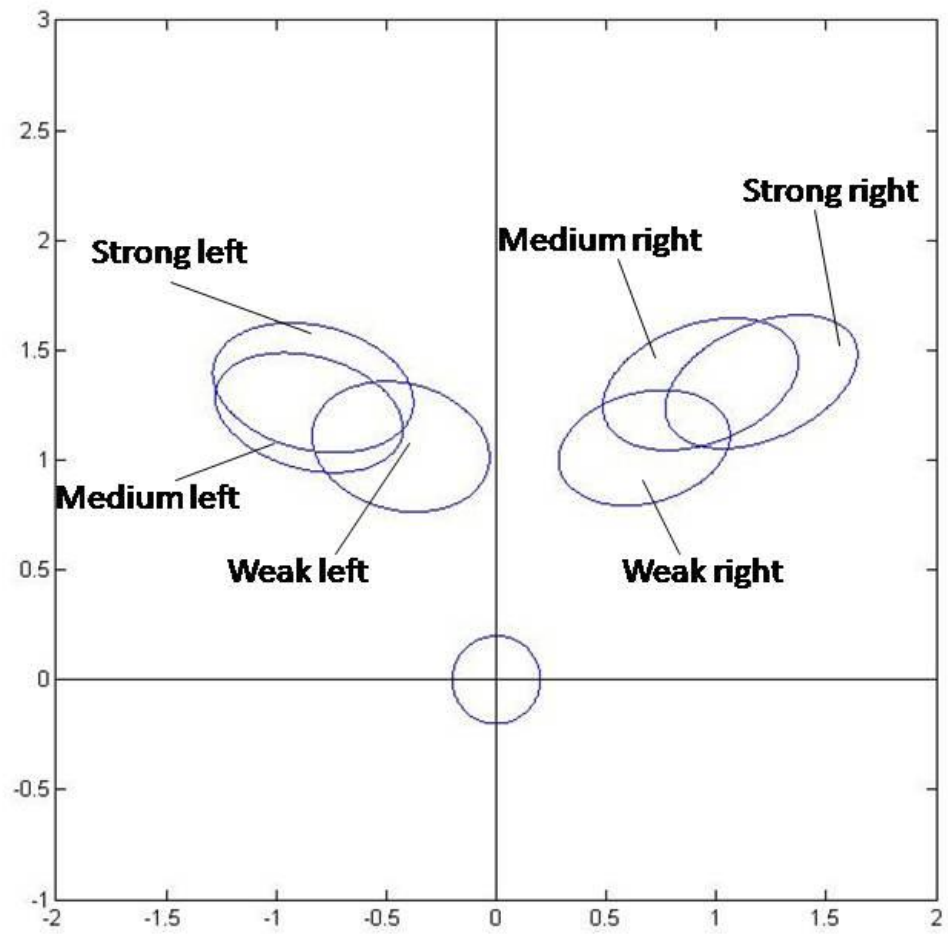
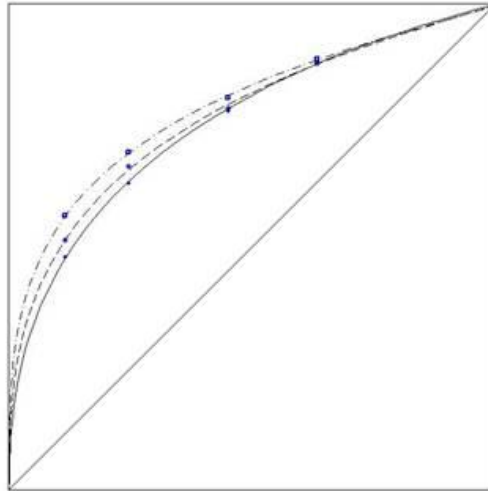




Figure 8

## Recognition



## Source Memory

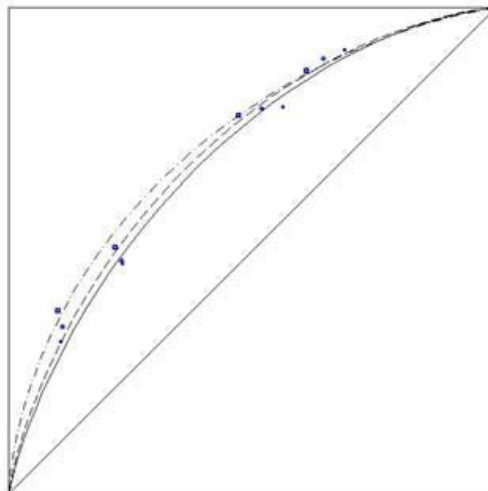
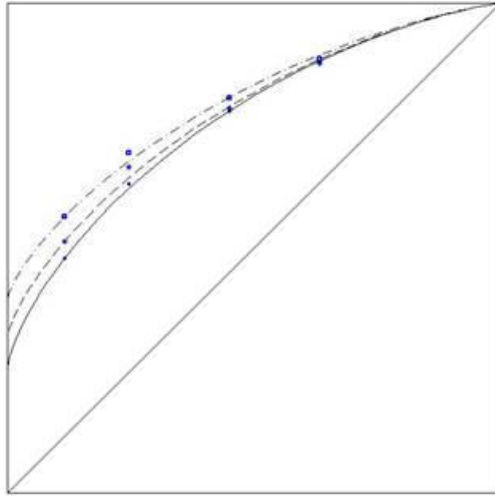


Figure 9

## Recognition



## Source Memory

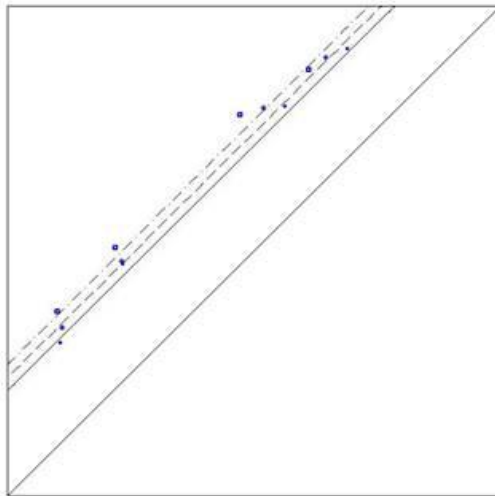
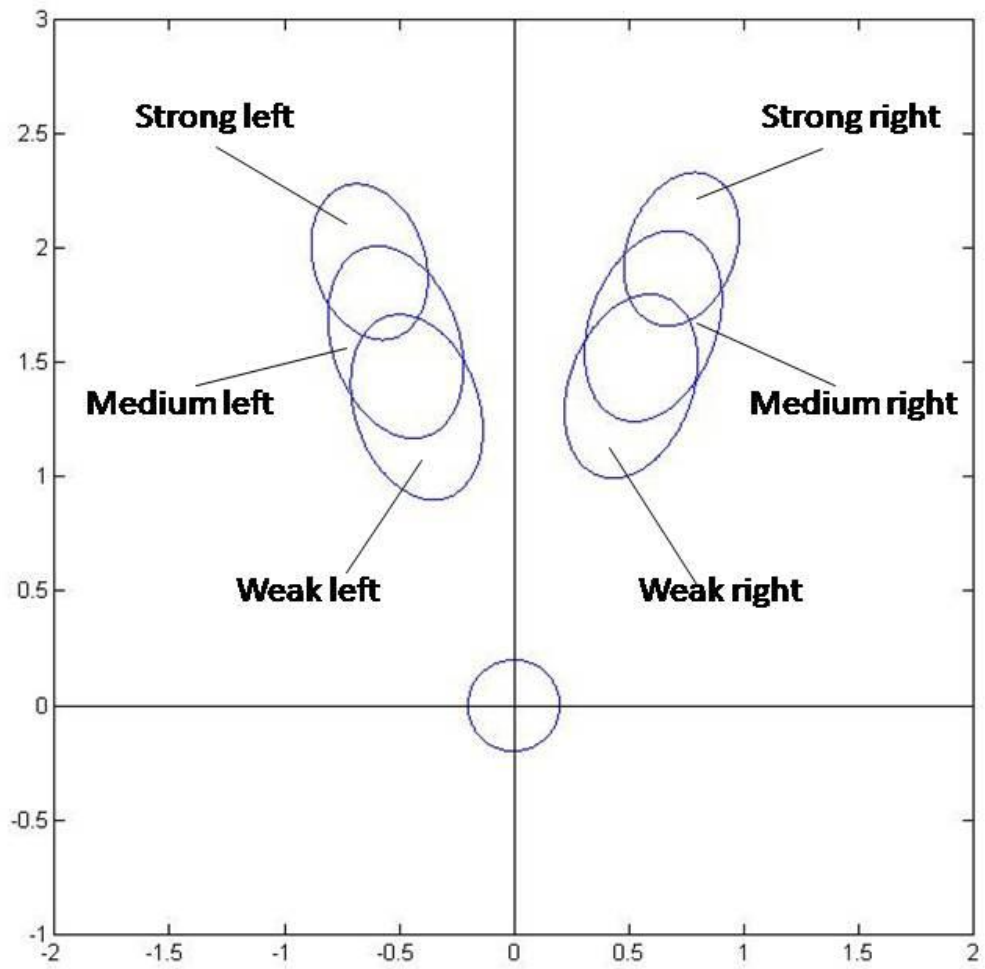


Figure 10



## Appendix

All models were fit by maximum likelihood estimation. Each model predicted the proportion of each response on the confidence scale for each item type, given a set of parameters. The multinomial likelihood function was then used to calculate the likelihood of these predicted proportions given the data (i.e., the actual frequency of each response). Specifically the multinomial likelihood function is:

$$L_w = \frac{N_w!}{\prod_{all\ r} \prod_{all\ s} x_{rsw}!} \prod_{all\ r} \prod_{all\ s} \theta_{rsw}^{x_{rsw}}$$

Where  $L_w$  is the likelihood for the  $w^{th}$  word type,  $\theta_{rsw}$  is the predicted proportion of words of the  $w^{th}$  word type getting both the  $r^{th}$  rating on the recognition test and the  $s^{th}$  rating on the source memory test.  $X_{rsw}$  is the actual frequency of the  $r^{th}$  and the  $s^{th}$  rating for the  $w^{th}$  word type. And  $N_w$  is the total number of items there were of the  $w^{th}$  word type. The joint likelihood of a model fit to all the different word types (i.e., studied-left and studied-right for each duration and new items) is the product of the likelihoods for each of those word types.

$$L_{joint} = \prod_{all\ w} L_w = \frac{\prod_{all\ w} N_w!}{\prod_{all\ w} \prod_{all\ r} \prod_{all\ s} x_{rsw}!} \prod_{all\ w} \prod_{all\ r} \prod_{all\ s} \theta_{rsw}^{x_{rsw}} \quad A1$$

### *Bivariate Model*

The bivariate model predicts the proportion of a response,  $\theta_{rsw}$ , is the volume under a bivariate curve that is bound by a square area that is defined by an upper and lower criterion on each of the two dimensions. Specifically:

$$\theta_{rsw} = \int_{c_{r-1}}^{c_r} \int_{c_{s-1}}^{c_s} N(\mu_w, \Sigma_w, x, y) dx dy \quad A2$$

where  $N(\mu_w, \Sigma_w, x, y)$  is a normal distribution with mean vector  $\mu_w$  and variance-covariance matrix  $\Sigma_w$ . The integrals are calculated with respect to both dimensions,  $x$  and  $y$ , and are bounded by the criteria  $c_{r-1}$  and  $c_r$  on the recognition dimension and  $c_{s-1}$  and  $c_s$  on the source dimension.  $\underline{R}$  and  $s$  both range from 1 to  $N_{\text{ratings}}$  and  $c_0 = -\infty$  and  $c_{N_{\text{ratings}}} = +\infty$ .

Given the parameters  $\mu_w$  and  $\Sigma_w$  for all  $w$ , the likelihood of a participant's data can be calculated by computing all predicted proportions,  $\theta_{rs w}$ , using equation A2 and then calculating the likelihood using equation A1. The set of parameters that maximized the likelihood for a given data set was found using the mle function in MATLAB, which instantiates a version of the Simplex search algorithm.

Each  $\mu_w$  vector consists of two parameters and each  $\Sigma_w$  consists of three parameters. In order to measure (and constrain) the angle the  $\mu_w$  vector makes with the recognition axes, the means on both dimensions were defined in polar coordinates. This allowed for the  $\mu_w$  vector to be defined by an angle,  $\gamma_w$ , and a length,  $r_w$ . The vector  $\mu_w$  then becomes:

$$\mu_w = \begin{bmatrix} r_w \cos \gamma_w \\ r_w \sin \gamma_w \end{bmatrix}$$

where  $r_w \cos(\gamma_w)$  is the mean on the recognition dimension for the  $w^{\text{th}}$  word type and  $r_w \sin(\gamma_w)$  is the mean on the source dimension for that word type.

Likewise, the variance-covariance matrix  $\Sigma_w$  was defined by its spectral decomposition (i.e., by decomposing it into two eigenvalues and two eigenvectors). Since a single angle is sufficient to define both eigenvectors, this allowed for the variance-covariance matrix to be defined in terms of two eigenvalues,  $\lambda_{w1}$  and  $\lambda_{w2}$ , and

the angle,  $\theta_w$ , that the major axis of that distribution's ellipse forms with the recognition axis. Specifically:

$$\Sigma_w = \lambda_{w1} e_{w1} e'_{w1} + \lambda_{w2} e_{w2} e'_{w2}$$

where  $\lambda_{w1}$  is the larger of the two eigenvalues and the eigenvectors,  $e_{w1}$  and  $e_{w2}$ , are of unit length and defined by the parameter  $\theta_w$  as follows:

$$e_{w1} = \begin{bmatrix} \cos \theta_w \\ \sin \theta_w \end{bmatrix} \quad \text{and} \quad e_{w2} = \begin{bmatrix} \cos(\theta_w + \frac{\pi}{2}) \\ \sin(\theta_w + \frac{\pi}{2}) \end{bmatrix}$$

$e_{w1}$  and  $e_{w2}$  also represent the direction of the major and minor axes of the ellipse for  $w^{\text{th}}$  word type, respectively. The length of those axes is proportional to the eigenvalues that correspond to those eigenvectors (and inversely proportional to the height of the iso-likelihood plane that was used to form the iso-likelihood contour plot).

#### *DPSD fit simultaneously to both tasks*

The DPSD model makes specific predictions for the recognition task and the source memory task independently. Specifically, the probability of endorsing an item on the recognition task is:

$$P(X > c_r \mid \text{studied-left}) = \underline{R} + (1 - R) * \Phi(d - c_r)$$

$$P(X > c_r \mid \text{studied-right}) = \underline{R} + (1 - R) * \Phi(d - c_r)$$

$$P(X > c_r \mid \text{new}) = \Phi(-c_r)$$

where  $X$  is the response variable,  $c_r$  is the  $r^{\text{th}}$  criterion,  $\Phi$  is the cumulative normal distribution,  $\underline{R}$  is the probability of an item being recollected, and  $\underline{d}$  is the distance between the means of the old distributions and the mean of the new distribution. The probability of endorsing an item on the source task is:

$$P(X > c_s \mid \text{studied-left}) = (1 - R) * \Phi(d - c_s)$$

$$P(X > c_s | \text{studied-right}) = \underline{R} + (1 - R) * \Phi(d - c_s)$$

$$P(X > c_s | \text{new}) = \Phi(-c_s)$$

where the scale is arbitrarily defined as higher values indicating more evidence for “rightness” and lower values indicating more evidence for “leftness.” Using the equations, it is possible to calculate the predicted proportion of each response on each test independently.

$$\theta_{r \bullet w} = P(X > c_r | w) - P(X > c_{r-1} | w)$$

$$\theta_{\bullet s w} = P(X > c_s | w) - P(X > c_{s-1} | w)$$

where  $\theta_{r \bullet w}$  is the proportion of the  $w^{\text{th}}$  word type given the  $r^{\text{th}}$  rating on the recognition test collapsing across all responses on the source memory test. Similarly,  $\theta_{\bullet s w}$  is the proportion of the  $w^{\text{th}}$  word type given the  $s^{\text{th}}$  rating on the source memory test collapsing across all responses on the recognition test.

However, in order to use the AIC statistic to compare models, the models must be fit to the same data. Some assumption then had to be made to get the DPSD model to be able to predict the proportion,  $\theta_{rsw}$ , of responses of the  $r^{\text{th}}$  rating on the recognition test and the  $s^{\text{th}}$  rating on the source test simultaneously. It was assumed that the proportion of responses on one test were statistically independent of the proportion of responses on the other test<sup>4</sup>, such that:

$$\theta_{rsw} = \theta_{r \bullet w} \theta_{\bullet s w}$$

---

<sup>4</sup> A different assumption would be to assume that recollected items always are always given the highest confidence correct rating on both tests (i.e., all  $\underline{R}$  responses go into a single cell in the 5 (recognition) X 5(source) response matrix for each word type) was also fit to the data from experiment 1. This assumption yielded markedly worse fits than the model that was used.

As with the bivariate model, given a set of parameters the likelihood for a participant's data was calculated by computing all predicted proportions,  $\theta_{rsw}$ , using equation A2 and then calculating the likelihood using equation A1. The parameters for the DPSD model are  $\underline{R}$  and  $\underline{d}$  for the different word- and test-types and the criteria on each test.