Toward a complete decision model
of item and source recognition

MICHAEL J. HAUTUS
University of Auckland, Auckland, New Zealand

AND

NEIL A. MACMILLAN AND CAREN M. ROTELLO
University of Massachusetts, Amherst, Massachusetts

In a recognition memory test, subjects may be asked to decide whether a test item is old or new (*item recognition*) or to decide among alternative sources from which it might have been drawn for study (*source recognition*). Confidence-rating-based receiver operating characteristic (ROC) curves for these superficially similar tasks are quite different, leading to the inference of correspondingly different decision processes. A complete account of source and item recognition will require a single model that can be fit to the entire data set. We postulated a detection-theoretic decision space whose dimensions, in the style of Banks (2000), are item strength and the relative strengths of the two sources. A model that assumes decision boundaries of constant likelihood ratios, source guessing for unrecognized items, and nonoptimal allocation of attention can account for data from three canonical data sets without assuming any processes specifically devoted to recollection. Observed and predicted ROCs for one of these data sets are given in the article, and ROCs for the other two may be downloaded from the Psychonomic Society's Archive of Norms, Stimuli, and Data, www.psychonomic.org/archive.

Recognition of episodic memories divides naturally into two parts: recognition that a test item has been encountered previously, and assessment of the source of that memory. The importance of source memory is obvious in many applications: The familiarity of a crime suspect has very different implications, depending on whether the memory arose from a crime scene or a chance encounter in a store; the potency of a memory of childhood abuse depends on whether the source is an early event or later discussions of the possibility of such an event.

Typical item and source experiments are parallel in structure. In item recognition, subjects study a list (of words, pictures, etc.) and are then tested with probes that may be targets (from the study list) or lures (unstudied). Each test probe is identified as "old" or "new," and confidence ratings may be added. In source recognition, subjects study two (or more) lists and are tested with probes that may be from either list. Each test item is identified as coming from List 1 or List 2, and confidence ratings may again be added.

Item and source recognition have both been studied using signal detection theory (SDT) techniques and models. Insight into the memory representation and decision processes can be obtained from receiver operating characteristic (ROC) curves, constructed by treating confidence ratings as different levels of response bias. In item recognition, correct responses to targets, called *hits*, are plotted against incorrect responses to lures, called *false alarms*. In source recognition, correct responses to items from one list, considered to be hits, are plotted against incorrect responses to items from the other list, considered to be false alarms.

In this article, we first summarize ROC data drawn from both item and source recognition. On the basis of the quite different ROC shapes these two tasks generate, theorists have inferred distinct underlying processes for each of them. We argue that such separatist modeling distorts conclusions and that a unified model of item and source recognition is needed. Two such models, both based on two-dimensional SDT, have been proposed, but these do not provide a good description of the existing data. We build stepwise on these models, adding likelihood-based criteria, source guessing for unrecognized items, and a degree of inattention. The final model provides an excellent fit to three canonical data sets.

MODELS OF SOURCE MEMORY ALONE

Form of Item and Source ROCs

To appreciate the distinctive characteristics of source ROCs, it is helpful to contrast them with their item recognition (old–new) counterparts. A representative item ROC is plotted on both probability and *z*-transformed coordinates in Figure 1. Many, many studies have replicated the result shown here: ROCs are curvilinear in probability space and linear in *z*-space. The slope of the *z*ROC is

M. J. Hautus, m.hautus@auckland.ac.nz

less than 1, averaging about 0.8. An SDT representation consistent with this outcome contains Gaussian target and lure distributions, the target distribution having a larger variance. Wixted (2007) argued that the unequal variances could arise because targets receive additional activation (by being presented on the study list) that varies across items. The continuous distributions imply a continuous strength variable on which decisions are based.

Source recognition ROCs are different. The first source ROCs were published by Yonelinas (1999), who collected the data to test an extension of his dual-process model of memory. In this model, source memory depends on both familiarity (a Gaussian process) and recollection (a high-threshold process). If sources are equated in familiarity, only recollection supports source memory, and the predicted ROC is linear; if familiarity makes a minor contribution, the ROC exhibits slight curvature. ROCs from three experiments that attempted to equate familiarity were slightly curved, confirming this prediction; the data from Experiment 2 are shown in Figure 2A. In a final experiment, all stimuli from one source were presented before stimuli from the other source, so that familiarity also clearly distinguished them. In that case, the ROC reverted to a curvilinear form consistent with either the dual-process model or the standard unequal-variance SDT model described earlier.

That source ROCs are not always linear, even when the role of familiarity is minimized, was demonstrated by Qin, Raye, Johnson, and Mitchell (2001). Their stimuli were statements by one of two people viewed on videotape. The resulting ROCs were curvilinear (as shown in Figure 2B), and although no fits were evaluated, the zROCs were roughly linear, as would be expected if the underlying distributions were Gaussian. Qin et al. argued that this result was consistent with the *source monitoring framework* (SMF; Johnson, Hashtroudi, & Lindsay, 1993), a generalization of Johnson and Raye's (1981) theory of reality monitoring that has guided much of the research in this area. According to the SMF, multiple features define a source, and different subsets of these features may determine source decisions; the complex stimuli of Qin et al. presumably provided many types of information about their sources. Yonelinas (1999) speculated that in his experiments a single detail distinguished the sources, whereas for Qin et al. the aggregation of items, even if discrete, led to effectively continuous distributions.

Slotnick, Klein, Dodson, and Shimamura (2000), also seeking to choose between continuous and threshold models by examining the curvatures of ROCs, used a standard male–female voice source paradigm. Their ROCs, although curved, were less so than a Gaussian model would predict; this third pattern to account for is illustrated in Figure 2C. Hilford, Glanzer, Kim, and DeCarlo (2002) replicated this result, as did Wong (2002) and DeCarlo (2003a). For ease of reference, we call this pattern a "flattened" ROC—that is, a curve more nearly linear than a pure SDT curve would be. Flattened ROCs turn out to have important theoretical implications.

Some insight into this range of results can be found in studies that have collected both old–new and source rat-
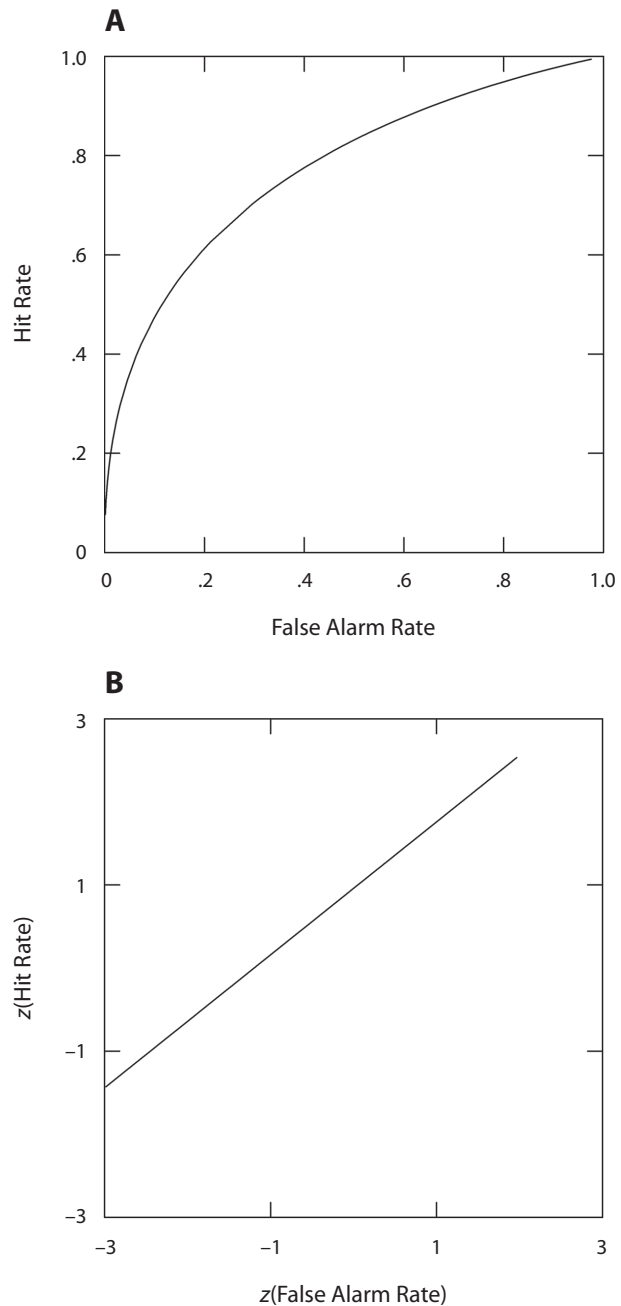

Figure 1. ROCs for the standard signal detection model are curved in probability space (A) and linear in z-space (B). The z-slope (typically 0.8 in old–new recognition) is interpreted as the ratio of the standard deviations of the underlying distributions.

ings. Taking advantage of this structure in their own and Yonelinas's (1999) study, Slotnick et al. (2000) were able to measure source ROCs either pooled across old–new confidence levels or contingent on a particular old–new rating. The "refined" ROCs based only on the highest-confidence "old" judgments appeared consistent with a Gaussian model, whereas the pooled ROCs were flatter. Slotnick and Dodson (2005) extended this analysis by plotting "refined" source ROCs contingent on each
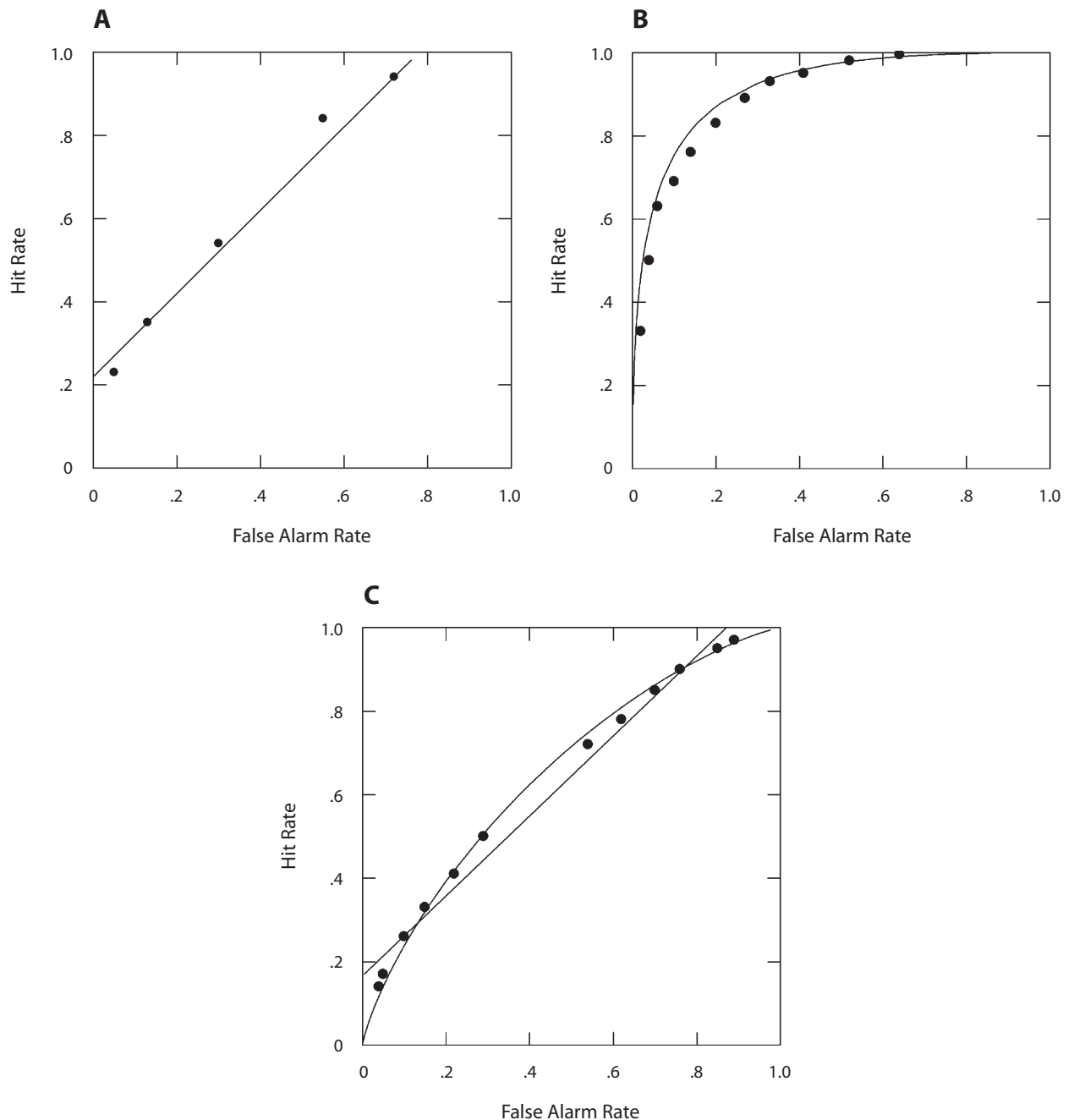
Figure 2. (A) ROC for source memory reported by Yonelinas (1999, Figure 4). Fits are for the dual-process model; notice that this model predicts a linear ROC for source memory. Adapted from "The Contribution of Recollection and Familiarity to Recognition and Source-Memory Judgments: A Formal Dual-Process Model and an Analysis of Receiver Operating Characteristics," by A. P. Yonelinas, 1999, *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *25*, p. 1425. Copyright 1999 by the American Psychological Association. (B) ROC for source memory reported by Qin et al. (2001, Figure 3), with points estimated using DataThief. Adapted from "Source ROCs Are (Typically) Curvilinear: Comment on Yonelinas (1999)," by J. Qin, C. L. Raye, M. K. Johnson, & K. J. Mitchell, 2001, *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *27*, p. 1113. Copyright 2001 by the American Psychological Association. (C) ROC for source memory reported by Slotnick et al. (2000, Figure 5). Fits are for the threshold and continuous (signal detection) models. Adapted from "An Analysis of Signal Detection and Threshold Models of Source Memory," by S. D. Slotnick, S. A. Klein, C. S. Dodson, & A. P. Shimamura, 2000, *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *26*, p. 1506. Copyright 2000 by the American Psychological Association.

level of old–new confidence. At the higher levels of "old" confidence, the refined ROCs were consistent with a Gaussian model, but they became increasingly flat as

confidence decreased. The intermediate level of curvature in the pooled ROCs noted by Slotnick et al., and replicated by Slotnick and Dodson, can thus be understood as

the result of averaging the source ROC data across high and low confidence that the items were studied. Slotnick and Dodson suggested that flatness of the refined source ROCs at low old–new confidence resulted from the addition of noise to the judgment. To determine which refined curves were contaminated by noise in this way, they calculated source d' values for each ROC and argued that cases in which d' was not significantly greater than 0 must have reflected the noise process. Reanalyzing the data of Yonelinas (1999), they found that the lowest four (out of six) ratings contributed noise by this definition. Slotnick and Dodson's own experiments manipulated exposure time; this variable had no effect on the form of the source ROC, but the model fits supported their noise model over multinomial and dual-process models.

The shift in ROC shape with old–new ratings provides another potential explanation of the discrepant ROC shapes reported by Yonelinas (1999) and Qin et al. (2001): Performance in the latter study was better (d' averaged 1.56, vs. 1.05 for Yonelinas, 1999), and accuracy levels were presumably related to confidence. Had old–new confidence judgments been taken, perhaps the Qin et al. curves would have reflected high-confidence refined curves, whereas the Yonelinas (1999) curves would have reflected low confidence. However, Qin et al. examined the shapes of ROCs for individual subjects and found curvature across all levels of performance. They did not report whether the ROCs were well-fit by SDT or had a lesser (or greater) degree of curvature.[1]

Research Strategies: A Limited Menu of Models and Analyses

The common attitude toward the data revealed in these studies is that the investigator's task is to choose between standard decision models. The titles of the articles summarized so far illustrate this implicit demand characteristic: "A Formal Dual-Process Model" (Yonelinas, 1999), "Signal Detection and Threshold Models" (Slotnick et al., 2000), and "A Continuous (Single-Process) Model" (Slotnick & Dodson, 2005). Published comments related to multinomial modeling (Kinchla, 1994) and dual-process modeling (Qin et al., 2001) have also contrasted off-the-shelf threshold and SDT models. These limited-menu decisions sometimes oblige authors to decide in favor of different models for different experiments. For example, Slotnick et al. (2000) concluded that "the results from Experiment 1 and Experiment 3 showed evidence in favor of the two-high threshold model and against the continuous model, whereas results from Experiment 2 showed evidence in favor of the continuous model and against the two-high threshold model" (p. 1513). If a single model is to account for all of the data we have described, it therefore will not be from the short list considered so far.

It is worth contrasting two statistical methods used to evaluate these models. First, each model predicts linearity in some form (the threshold model for ROCs in probability space, the SDT model for zROCs) and nonlinearity in another (the SDT model for ROCs in probability space, the threshold model for zROCs). To test these predictions, some investigators (e.g., Glanzer, Hilford,

& Kim, 2004) have fit quadratic functions to the data, rejecting the linearity hypothesis if the quadratic term is reliably greater than 0. Thus, for example, SDT theorists point to significant curvature in the source ROC and a lack of curvature in the zROC for item recognition (Hilford et al., 2002; Wixted, 2007). Weaknesses in this approach, acknowledged even by some of those who use it (Slotnick & Dodson, 2005), are that no model predicts a quadratic form for the ROC and that not all curvatures are equivalent. An example of the pluses and minuses of linearity analysis can be found in a recent article by Parks and Yonelinas (2007), who examined 59 ROCs from the source, associative-recognition, and plurals paradigms. Of the source zROCs, 27 of 30 were fit best by quadratics with coefficients nominally greater than 0. This finding is contradictory to a simple SDT model, as they note, but contrary to their claim, it does not specifically support a dual-process theory: Nonlinear source zROCs are also predicted by all of the new models to be tested in this article, and none of them include a threshold component.

The second statistical method, used in this article, is to compare goodness-of-fit measures for the models' ability to describe the entire set of data. The most common index of fit, the sum of squared errors, is inappropriate (since both axes in ROC space are variable and successive operating points are not independent), but χ^2 evaluation of maximum-likelihood estimation (MLE) fits is sometimes performed (Slotnick et al., 2000). We will ourselves use MLE and related methods in the evaluation of our models. Relying on goodness of fit has its own shortcomings: A model may achieve good absolute fits simply because it is highly flexible and complex (see, e.g., Cohen, Rotello, & Macmillan, 2008). Where possible, we reduce this concern in what follows by comparing nested models; we also report Akaike information criterion (AIC; Akaike, 1973) and Bayesian information criterion (BIC; Schwarz, 1978) measures that allow comparisons of nonnested models with different numbers of parameters. The AIC and BIC control one aspect of model complexity, the number of free parameters, but do not equate the models' functional forms (see Pitt, Myung, & Zhang, 2002). This concern is somewhat alleviated if the parameter estimates make substantive sense, because complex models that overfit the data (i.e., "fit noise") may require unrealistic parameter values to succeed.

MIXTURE MODELS

What could account for ROCs that are consistent neither with Gaussian models nor with those with threshold components? One possibility is that the underlying distributions have a different shape entirely, but in that case, the finding that the old–new ROCs collected in source experiments are well-described by Gaussian models would require explanation. Another explanation lies in the possibility that Gaussian strength distributions are consulted on most, but not all, trials; on the remaining trials, a different Gaussian distribution could be consulted. On this view, the responses reflect a mixture of these underlying distributions.

DeCarlo (2002) proposed a "mixture distribution" extension of SDT, focusing on applications in recognition memory. Figure 3A illustrates the basic premise: Testing an item that was presented leads to a sample from one of two memory distributions, A or A', that differ in means but not in variance. In one interpretation, the alternative distributions arise from intermittent attention: The A distribution reflects cases for which the studied item was attended; A' reflects cases for which it was not. The A' distribution may or may not have a mean equal to that of the lure distribution. For many parameter values, this mixture model predicts ROCs consistent with the standard unequal-variance SDT model, because the mixture of the two target distributions with different means and the same variance has a greater variance than either of them alone. The resulting zROC, which is approximately linear, has a slope less than 1 because mixing occurs only for target trials. DeCarlo (2002) fit classic data from recognition experiments in which word frequency and exposure duration (which presumably affects attention) were manipulated; he applied both the mixture and unequal-variance SDT models. These models were about equally successful with the group ROC data, but the mixture model was better able to fit a variety of unusual ROC shapes produced by the individual subjects. In addition, DeCarlo (2002) argued that interpreting the A' distribution as inattention provides a rationale for the model more satisfactory than any that has been proposed for the unequal-variance model.

In the source memory domain, the mixture model is intuitively compelling: Subjects with limited attention may encode features relevant to item content that do not serve to identify the source. DeCarlo (2003a) then extended the model to source judgments; a version in which only studied items are tested is illustrated in Figure 3B. The unattended distribution (U) now lies almost midway between the two target distributions, so judgments of items from sources A and B are both based on the mixture distributions A-or-U and B-or-U. Because both source strengths are mixtures and no lures are tested, the resulting zROC is generally concave-upward and the ROC is flattened, consistent with empirical observations. Like its item recognition cousin, however, the source mixture model is flexible, generating zROCs with a variety of shapes, depending on subjects' sensitivity and attention to the two sources. De-Carlo (2003a) showed that the model quantitatively fits the source data of Yonelinas (1999), as well as those of Hilford et al. (2002); it is also at least qualitatively consistent with the data of Slotnick et al. (2000).

Slotnick and Dodson's (2005) account of source ROC shapes can be understood in terms of a mixture model. Items receiving high-confidence "old" judgments are those that were fully attended: For those items, the probability of sampling a strength from the U distribution would be essentially 0. As a result, the distributions determining the refined ROC are Gaussian, and the ROC is SDT-like. In contrast, items receiving lower ratings are those that were attended on a fraction of trials, have distributions underlying their refined ROCs that are mixtures, and have ROCs with lesser degrees of curvature. Slotnick
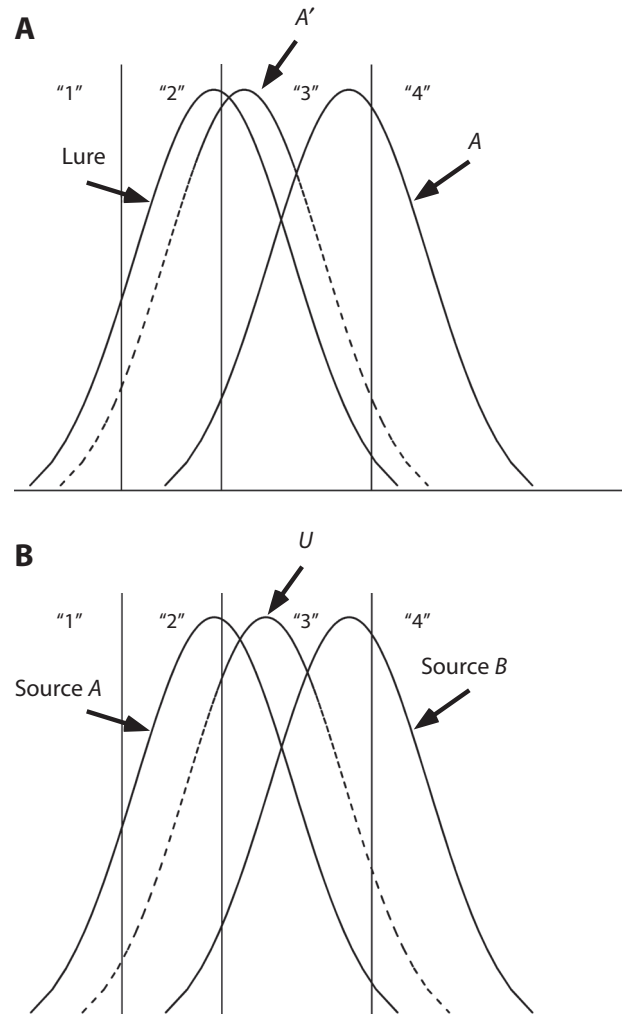

Figure 3. DeCarlo's (2002) mixture model. (A) Recognition: On lure trials, a sample from the "Lure" distribution is presented. On target trials, a sample from A occurs with probability λ, and a sample from A' occurs with probability $1 - \lambda$. Three criteria divide the decision axis into regions that correspond to different levels of confidence. Adapted from "Signal Detection Theory With Finite Mixture Distributions: Theoretical Developments With Applications to Recognition Memory," by L. T. DeCarlo, 2002, *Psychological Review*, *109*, Figure 1, p. 711. Copyright 2002 by the American Psychological Association. (B) Source memory: A sample from either source A or source B is presented. With probability λ, a sample is drawn from the relevant distribution A or B, but on a proportion $1 - \lambda$ of trials, the subject is inattentive, and the sample is drawn instead from U. The effective distributions for both stimulus classes are mixtures. Adapted from "An Application of Signal Detection Theory With Finite Mixture Distributions to Source Discrimination," by L. T. DeCarlo, 2003, *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *29*, Figure 1, p. 768. Copyright 2003 by the American Psychological Association.

and Dodson's "added noise" proposal can also be seen as an interpretation of the mixture process.[2] Mixture models broaden the range of ROC shapes that can be accounted for, provided that they can be incorporated in a model detailed enough to describe all of the data. The next step is to develop a framework for such a model.

MODELS THAT INCLUDE BOTH ITEM AND SOURCE RECOGNITION

In some experiments and real-life situations, the only decision to make is one of source, but there are probably more circumstances in which both item and source judgments are required. "Have you seen this person before?" (item recognition), asks the prosecutor, and "What was the occasion?" (source judgment). So far we have been modeling one decision at a time, but this approach is logically inconsistent if (as we assume) subjects are building a representation during study, aspects of which are used at test to support both item and source judgments. In this view, a decision space from which item and source judgments both arise is necessary, and the remainder of this article considers models on the basis of such representations.

The simplest SDT model for three stimulus classes is one-dimensional. Figure 4 illustrates an arrangement in which each of the two sources has greater average strength than the lures. Source d' is predicted to equal the difference between the two old–new recognition d' values. A second possibility, not shown, is that items from one source average greater strength than the lures, but those from the other source average less strength. In that case, source d' is predicted to equal the sum of the two item d' values. Empirically, however, observed source d' usually falls between these two possibilities. For example, Glanzer et al. (2004) estimated both item d' and source d' in five experiments, all of which employed lists of words as stimuli. In Experiment 1, the item d' values were 1.13 and 1.14, so a representation like Figure 4 would imply source $d' = 1.14 - 1.13 = 0.01$. On the other hand, if items from different sources differed from the lures in op-

posite directions, source d' would be $1.13 + 1.14 = 2.27$. Neither possibility was supported: The source d' equaled 1.10. Glanzer et al.'s other four experiments showed the same pattern.

One experiment has seemed to show support for a one-dimensional model. In Hoffman's (1997) reality-monitoring study, subjects either saw or imagined items in sessions 2 days apart, receiving a source ("real" or "imagined") test following the second study list. More accurate recognition judgments were obtained for whichever type of item was encountered more recently, and source judgments were fit by a one-dimensional SDT model with apparent success. This finding appears not to be general, however. Bink, Marsh, and Hicks (1999) replicated Hoffman's experiment but strengthened the Day 1 items so as to equate performance on the two sources. The predicted source accuracy ($d' = 0$) was not obtained; rather, the proportion correct for those judgments was about .90.

These findings are likely to be unsurprising to many. The one-dimensional model is in conflict with the SMF perspective that multiple features (possibly reflecting independent types of information) contribute to source discrimination, and with the intuition that the dimension on which sources differ is qualitatively different from memory strength. The remainder of this article considers only models in which separate dimensions of item and source information are made explicit.

The Banks (2000) Model: Uncorrelated Distributions and Linear Bounds

A natural improvement over the one-dimensional model locates the means of the three distributions in a two-dimensional space. Banks (2000) adapted a perceptual model by Tanner (1956) to the source memory problem. In his experiments, the stimuli were words and proper names, and the sources were visual and auditory presentation. He conducted item and source memory tests and found a pattern of sensitivities similar to that in the Glanzer et al. (2004) data discussed earlier. The data thus were poorly fit by a one-dimensional model, but they were well-described by a two-dimensional decision space, shown in simplified form in Figure 5. The dimensions are interpreted as overall strength (the vertical axis in the figure) and relative source strength (the horizontal axis). Each stimulus class generates a bivariate normal distribution in the space; circles in the figure are contours of equal likelihood for those distributions. Banks proposed that the two axes serve as the bases for recognition and source decisions: A horizontal line divides the space into regions corresponding to "old" and "new" responses, and a vertical line determines the source judgment.

The DeCarlo (2003b) Model: Correlated Distributions and Linear Bounds

Banks's (2000) model appears to predict SDT-like ROC curves. It requires modification to account for the ROC data we have discussed, particularly the changes in ROC shape and source accuracy as a function of old–new confidence found by Slotnick and Dodson (2005). De-Carlo (2003b) proposed that the distributions in the Banks
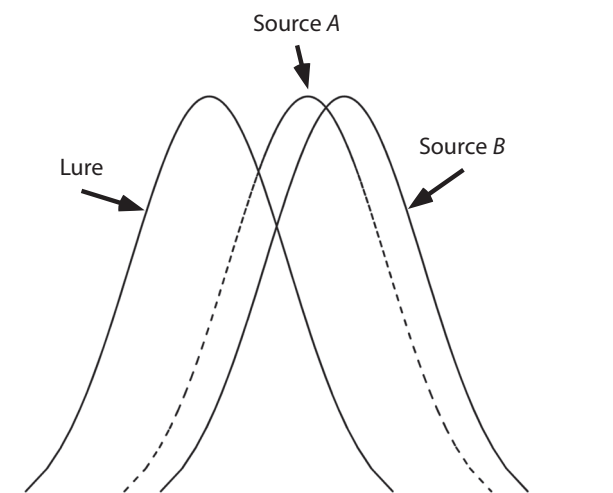
Figure 4. An implausible signal detection model for item and source recognition in which a single strength variable mediates all judgments. Source discrimination (between *A* and *B*) is poor as compared with item detection (between *A* and lures or between *B* and lures). If, on the other hand, the strength of *A* is imagined to be systematically lower than the strength of the lures, source discrimination is very high. In most data, source discrimination is similar in magnitude to item recognition, contrary to both versions of the single-strength model.
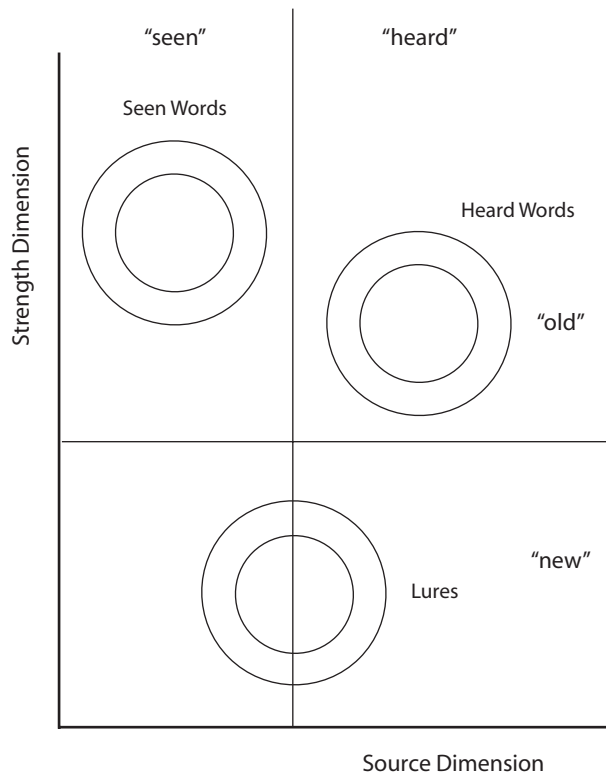
Figure 5. The decision space derived by Banks (2000) for old–new discrimination and source memory. The circles represent mean equal-likelihood contours for bivariate, equal-variance normal distributions of the stimulus classes. The vertical axis represents item strength, and the horizontal axis, the relative strengths of the two sources. Adapted from "Recognition and Source Memory As Multivariate Decision Processes," by W. P. Banks, 2000, *Psychological Science*, *11*, Figure 3, p. 270. Copyright 2000 by Blackwell Publishing, Inc.

(2000) model should have nonzero bivariate correlations, as shown in Figure 6. The correlations arise because some of the same information contributes to both source and old–new decisions: Subjects whose source judgments are highly accurate when item strengths are high show poorer source accuracy on test items that they are not even sure they studied. In DeCarlo's (2003b) proposal, as for Banks, the item and source decision boundaries between responses are linear and orthogonal to each other.

DeCarlo (2003b) pointed out that if this model is correct, the standard analyses, which treat source and item information separately, provide biased estimates of source memory: Accuracy appears to be higher when source judgments are only requested for "old" responses (reflecting greater memory strength) than when they are requested even for items the subject has claimed are "new" (reflecting lower strength). Valid measures of recognition and source sensitivity and other parameters can be obtained only by fitting a complete multidimensional model to the full set of data. Such a model must describe ROC curves of four types: pooled ROCs for item and for source recognition; refined ROCs for source judgments, contingent on old–new rating; and refined ROCs for item recognition, contingent on source rating. The data of Yonelinas (1999)

can be described with 19 parameters: the vertical distances of the source distributions (*A* and *B*) from the origin, the horizontal distance between the means of the source distributions, two standard deviations and a correlation for each of the source distributions, and five old–new and five source rating criterion locations.[3] Because Slotnick et al. (2000) used seven rating categories instead of six for both old–new and source decisions, two additional criterion location parameters are required to model their data.

DeCarlo (2003b) calculated such fits for the data of Yonelinas (1999, Experiment 2) and Slotnick et al. (2000, Experiments 2 and 3). We have generally followed DeCarlo's (2003b) strategy for the models proposed in the present article, so here we describe his methods in some detail. Fitting was accomplished by MLE, with no constraints placed on the parameters. Negative log likelihood was minimized using the downhill simplex method; the algorithm was restarted a number of times with different parameters in an effort to avoid local minima. We calculated two goodness-of-fit statistics: The first was $G^2 = 2\Sigma O \ln(O/\hat{E})$, where O and $\hat{E}$ are the observed and expected frequencies in each cell of the contingency table, and the sum is over all cells. G^2 is asymptotically distributed as χ^2, with degrees of freedom (*df*s) in the present context equal to $3(C_{\mathrm{ON}}C_{\mathrm{S}} - 1) - N_{\mathrm{p}}$, where C_{ON} and C_{S} are the numbers of old–new and source rating categories, respectively, and N_{p} is the number of free parameters in the model.

Complementary information is provided by the root mean square error of approximation (RMSEA). DeCarlo (2003b) reported this measure for the models that he fit-
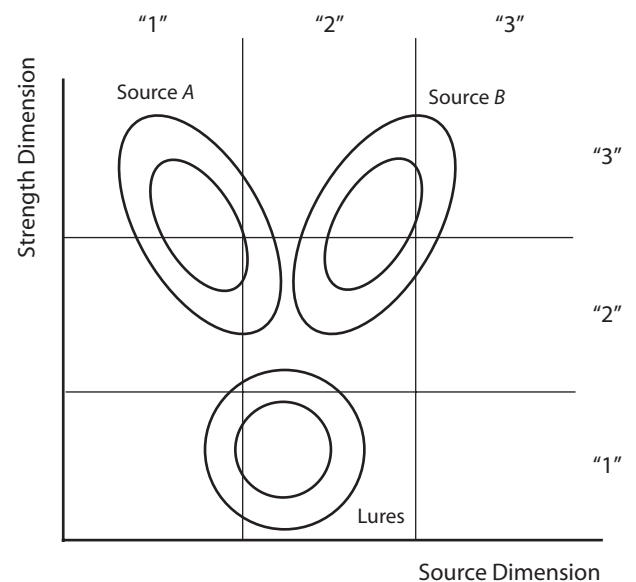

Figure 6. DeCarlo (2003b) correlated-distribution model. The distributions of two source items *A* and *B* display correlations between memory strength (*y*-axis) and source information (*x*-axis). As a result, source accuracy (the distance between *A* and *B*) is greater when item strength is high (upper region) than when it is low. From "Source Monitoring and Multivariate Signal Detection Theory, With a Model for Selection," by L. T. DeCarlo, 2003, *Journal of Mathematical Psychology*, *47*, Figure 1, p. 294. Copyright 2003 by Elsevier Ltd. Adapted with permission.

Table 1
Parameters of Models

Model	d_A	d_B	d_{AB}	$\sigma^2(xA)$	$\sigma^2(yA)$	$\sigma^2(xB)$	$\sigma^2(yB)$	r_A	r_B	d_{xN}	d_{yN}	$\sigma^2(xN)$	$\sigma^2(yN)$	p
					Yonelinas (1999, Experiment 2)									
DeCarlo	1.41	1.34	1.45	1.82	1.49	1.77	1.46	−.43	.48					
Model 1	1.39	1.30	0.83	1.06	1.47	1.07	1.47	−.23	.32					
Model 2	1.42	1.36	0.61	1.06	1.47	1.07	1.46	−.41	.47					
Model 3	3.75	3.76	1.23	1.12	2.23	1.07	2.50	−.42	.36	0.32	0.90	0.92	1.06	.46
					Slotnick et al. (2000, Experiment 2)									
DeCarlo	2.78	2.78	4.67	3.35	1.90	3.39	2.00	−.50	.60					
Model 1	3.08	3.10	2.25	1.92	2.24	1.67	2.12	−.50	.23					
Model 2	2.70	2.83	1.75	1.67	1.83	1.54	2.01	−.58	.38					
Model 3	5.25	5.41	3.99	2.81	2.72	2.71	2.91	−.51	.38	0.00	1.30	2.45	2.26	.18
					Slotnick et al. (2000, Experiment 3)									
DeCarlo	2.02	2.03	2.90	2.87	1.49	2.92	1.49	−.51	.52					
Model 1	2.11	2.12	1.58	1.57	1.59	1.56	1.62	−.26	.32					
Model 2	1.85	1.86	1.01	1.39	1.34	1.16	1.32	−.56	.23					
Model 3	2.91	2.92	1.70	1.49	1.48	1.39	1.42	−.44	.34	−0.18	1.02	1.18	1.12	.25

Note—"DeCarlo" indicates the DeCarlo (2003b) model. A and B are the two source distributions, and N is the noise distribution. σ^2 is variance on the x- and y-axes; d_A and d_B are the vertical distances between the target and lure distributions; d_{xN} and d_{yN} are the distances in the x and y directions, respectively, of the noise distribution mean from the origin; d_{AB} is source accuracy; and p is the proportion of nonattended trials.

ted, and the appendix to his article provides an introduction to it. In general, RMSEA < .05 indicates close fit, .05 < RMSEA < .08 acceptable fit, and RMSEA > .10 poor fit (Browne & Cudeck, 1993).

Our implementation of DeCarlo's (2003b) model differed slightly from his, so we repeated his calculations.[4] In Table 1, the rows labeled "DeCarlo" contain the parameter estimates, which were generally reasonable.[5] In particular, the bivariate correlations of the target distributions were between .4 and .6 in absolute value and (as in Figure 6) opposite in sign. Like DeCarlo (2003b), we found the fit of the model to all of the data sets to be poor. For purposes of allowing comparison with our own models, we calculated G^2 for each experiment, given the best-fitting DeCarlo (2003b) model. These values, reported in Table 2, differ only slightly from those reported by DeCarlo (2003b, Table 1), indicating that the different parameterization of the model we adopted had little effect on goodness of fit.

The model's ROC predictions are compared graphically with Yonelinas's (1999) data in Figure 7. The four smaller panels (top) display ROCs pooled across all ratings on the other task. The left pair illustrates the old–new pooled ROC and zROC, and the model provides a good fit. The right pair of panels shows the pooled source data, and here the model systematically fails. The predicted zROCs are linear (because of the normal distributions and linear bounds), whereas the data display curvature. The two larger panels (bottom) illustrate the refined old–new and source ROCs. There are six refined ROCs of each type, one for each level of confidence on the other task; for example, the six source ROCs are obtained from the six levels of old–new confidence.

There are two major observable patterns in the refined-ROC data. First, for the old–new ROCs, maximum sensitivity (as measured by the area under the curve) occurs for the extreme source ratings ("sure source A" and "sure source B"), and lower sensitivity is found for intermediate source ratings. Second, for the source ROCs, sensitivity is at a maximum for the most confident old–new rating and drops away rapidly for lower levels of confidence. The refined curves exhibit substantial discrepancies between the data and the model: The predicted old–new ROCs are too

Table 2
Fits of Four Models to Three Data Sets

	Yonelinas (1999, Experiment 2)				Slotnick et al. (2000, Experiment 2)				Slotnick et al. (2000, Experiment 3)			
Model	p/df	G^2	ΔG^2	RMSEA	p/df	G^2	ΔG^2	RMSEA	p/df	G^2	ΔG^2	RMSEA
DeCarlo (2003b)[a]: Linear source bounds	19/86	2,418	—	0.119	21/123	1,582	—	0.117	21/123	3,765	—	0.124
Model 1: LR source bounds	19/86	951	—	0.072	21/123	620	—	0.068	21/123	1,186	—	0.067
Model 2: LR source bounds for "old" responses, guessing for "new" responses	24/81	213	—	0.029	27/117	199	—	0.029	27/117	212	—	0.021
Model 3: Model 2 plus arbitrary inattentive noise	29/76	178	35	0.026	32/112	173	26	0.025	32/112	193	19	0.019

Note—p/df, number of parameters/degrees of freedom, not corrected for empty cells. Some researchers advocate removing one df for each empty cell; for consistency with past work, we have followed DeCarlo (2003b) in not adjusting the dfs in that way. G^2, statistic for goodness of fit (smaller numbers indicate better fit); ΔG^2 provides a χ^2 test of nested Models 2 and 3, with 5 df. RMSEA, a measure of approximate fit (RMSEA < .05, close fit; .05 < RMSEA < .08, acceptable fit; RMSEA > .10, poor fit). LR, likelihood ratio.   [a]The DeCarlo model is slightly modified for this study, so our values do not perfectly match.
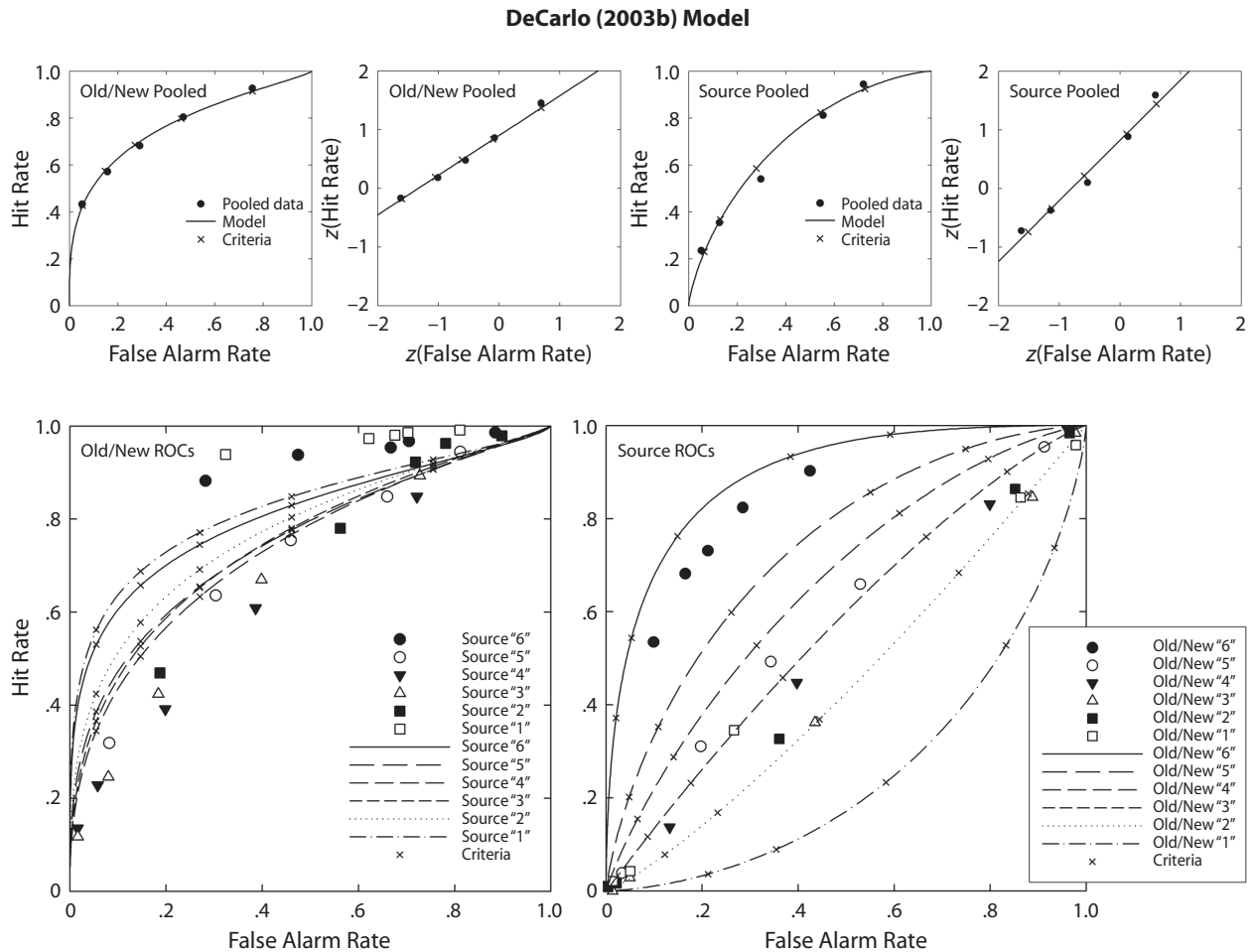
DeCarlo (2003b) Model

Figure 7. The DeCarlo (2003b) model fit to the ROC data of Yonelinas (1999, Experiment 2). The left half of the figure shows old–new data and fitted curves; the right half shows source data and curves. The upper panels report pooled analyses (in probability space and z-space), and the lower panels reflect refined analyses (see the text). Symbols indicate observed data, whereas the curves and ×s show the ROC and criterion-location predictions of the model.

compressed, and the refined source ROCs are too spread out. The two Slotnick et al. (2000) data sets met similar fates with this model (see Table 2).

To our knowledge, refined old–new ROCs have not previously been plotted for these (or any other) data, but they contain information that is just as valuable as refined source ROCs. For the present data sets, old–new ROCs display above-chance performance for all levels of source rating, whereas refined source ROCs fall rapidly to near chance when old–new confidence is less than maximal. This new way of representing the data allows additional insight into the discrepancies between the predicted and observed curves.

In the remainder of the article, we gradually improve this model, seeking a detailed picture of how old–new and source judgments are combined. Like DeCarlo (2003b), we examined the Yonelinas (1999) and Slotnick et al. (2000) data sets, which to our knowledge are still the only ones published with rating response frequencies for both old–new and source judgments.[6] For the sake of continu-

ity, the particular experiments we examined are the same as those analyzed by DeCarlo (2003b).

APPROXIMATIONS TO A COMPLETE RECOGNITION/SOURCE MODEL

The failure of DeCarlo's (2003b) model encouraged us to look at its assumptions more carefully. The models we consider here are all detection-theoretic variations on the DeCarlo (2003b) theme. They assume that the distributions for the lures and both targets are bivariate normal and that decisions are made by partitioning the space with fixed boundaries. In the modified models proposed below, three assumptions are relaxed: the linearity of the decision bounds, the assumption that subjects attend completely to items presented for study, and the assumption that source judgments are made in the same way following "new" and "old" recognition responses. The literature summarized earlier has provided the building blocks for a complete decision model for source memory data that provides an

excellent description of the data. By moving stepwise away from the DeCarlo (2003b) starting point, some insight into the processes underlying source memory can be obtained.

Model 1: Likelihood-Ratio Source Bounds

An important restriction of the DeCarlo (2003b) model, as illustrated in Figure 6, is that all decision boundaries are straight lines. In general recognition theory (Ashby & Townsend, 1986), such straight-line bounds parallel to the strength axes are said to display *decisional separability*, which might be viewed as a cognitively simple rule. For the criteria that determine the old–new judgment, linearity is a reasonable assumption. The effective decision axis becomes a line from the lure toward the target distributions, as proposed by Banks (2000; see Figure 5), and the model for this judgment reduces to a simple one-dimensional Gaussian one. The predicted old–new zROC is a straight line, consistent with most findings in the literature.

By the same line of reasoning, however, DeCarlo's (2003b) proposed source criteria are much less plausible. If these were indeed linear, source ROCs would have the same Gaussian-based shape as those for item recognition, and we have seen that this is rarely true. A look back at the fits of DeCarlo's (2003b) model (Figure 7) illuminates the consequences of the linearity assumption. The predicted refined source ROCs for old–new ratings of 1 and 2 are clearly below chance. This oddity results from the bivariate correlations, which guarantee that the two source distributions will gradually merge and then cross as old–new ratings decrease. In the lower region of the decision space (see Figure 6), the linear bounds force decisions that are exactly the opposite of the correct ones.

In Model 1, we replaced linear source bounds with rules based on likelihood ratio (LR), the relative likelihoods of the two events being discriminated. LR bounds are linear when only two stimulus classes are to be discriminated, and the resulting bivariate distributions have identical covariance matrices. LR bounds can be strongly nonlinear, however, when variances are unequal or the distributions contain unequal bivariate correlations, as in the DeCarlo (2003b) model. Curved LR bounds for the Yonelinas (1999) data are illustrated in Figure 8. Each is composed of two parts; for example, the highest rating in favor of the *A* source is given for observations in the upper-left region of the space, which is unsurprising, but also for those in the extreme lower-right region. The second component is due to the crossing of the two source distributions that arises from the bivariate correlations. The numbers in the diagram indicate regions for which ratings 6 through 1 are appropriate.[7]

The LR model accounts for the data better than the linear-criterion model; Table 2 shows that G^2 is smaller for all three data sets. Because the two models have the same number of dfs—with criterion locations along the axes replaced by values of likelihood ratio—this direct comparison of G^2s is appropriate.[8] The successes and failures of the model can be seen in Figure 9, where the LR model is fitted to Yonelinas's (1999) ROC data. The model provides a good fit to the pooled old–new data,
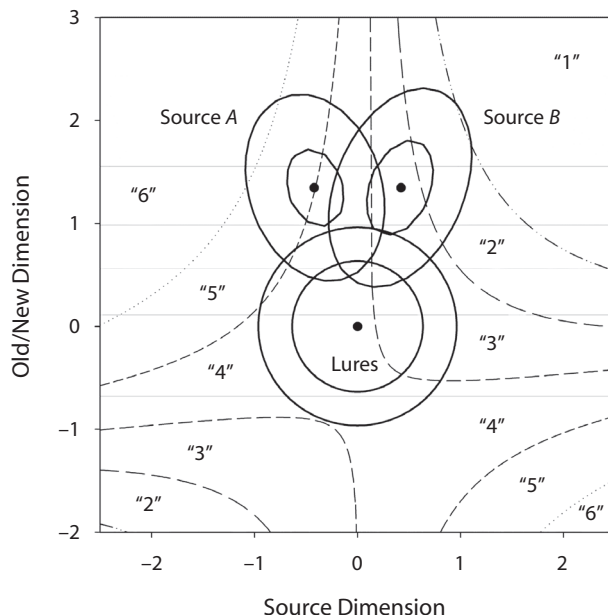

Figure 8. Two-dimensional decision space for recognition and source memory, with decision boundaries forming contours of constant likelihood ratio. The distributions and boundaries were obtained by applying Model 1 to the Yonelinas (1999) data. Each number is the source rating associated with each region (6 = "sure source *A*"; 1 = "sure source *B*").

and a relatively good fit to the pooled source data. In particular, it predicts a flattened source ROC in probability coordinates and a curvilinear zROC. These patterns are found in the data, but the flattening effect is not as great as the data require. The refined source ROCs capture the major patterns in the data—lower performance and flatter ROCs as old–new confidence declines—but the fits are not convincing.

Perhaps the most surprising outcome is the ability of the model to predict source ROCs that are curved upward in z-space. Such curvature has long been a signature prediction of threshold or mixture models, and has been considered inconsistent with SDT. For example, Yonelinas and Parks (2007) wrote, "The U-shaped zROCs that are consistently observed in relational recognition studies [such as source experiments] cannot be accounted for by a pure signal detection model" (p. 819). This statement is clearly incorrect for Model 1, which is "a pure signal detection model," albeit in two dimensions.

We also analyzed Experiments 2 and 3 of Slotnick et al. (2000), with similar results. One difference between their experiments and those of Yonelinas (1999), previously mentioned by DeCarlo (2003b), is the greater preponderance of 0s in the data matrices. This pattern results from the higher sensitivities in those experiments: Table 1 reveals that target d' values averaged 3.09 (Experiment 2) and 2.12 (Experiment 3), whereas in Yonelinas's (1999) experiment they averaged 1.35. Statistical parsing of Slotnick et al.'s (2000) data is thus more difficult. We do not present here figures for their experiments analogous to Figure 8, but do comment below when their findings
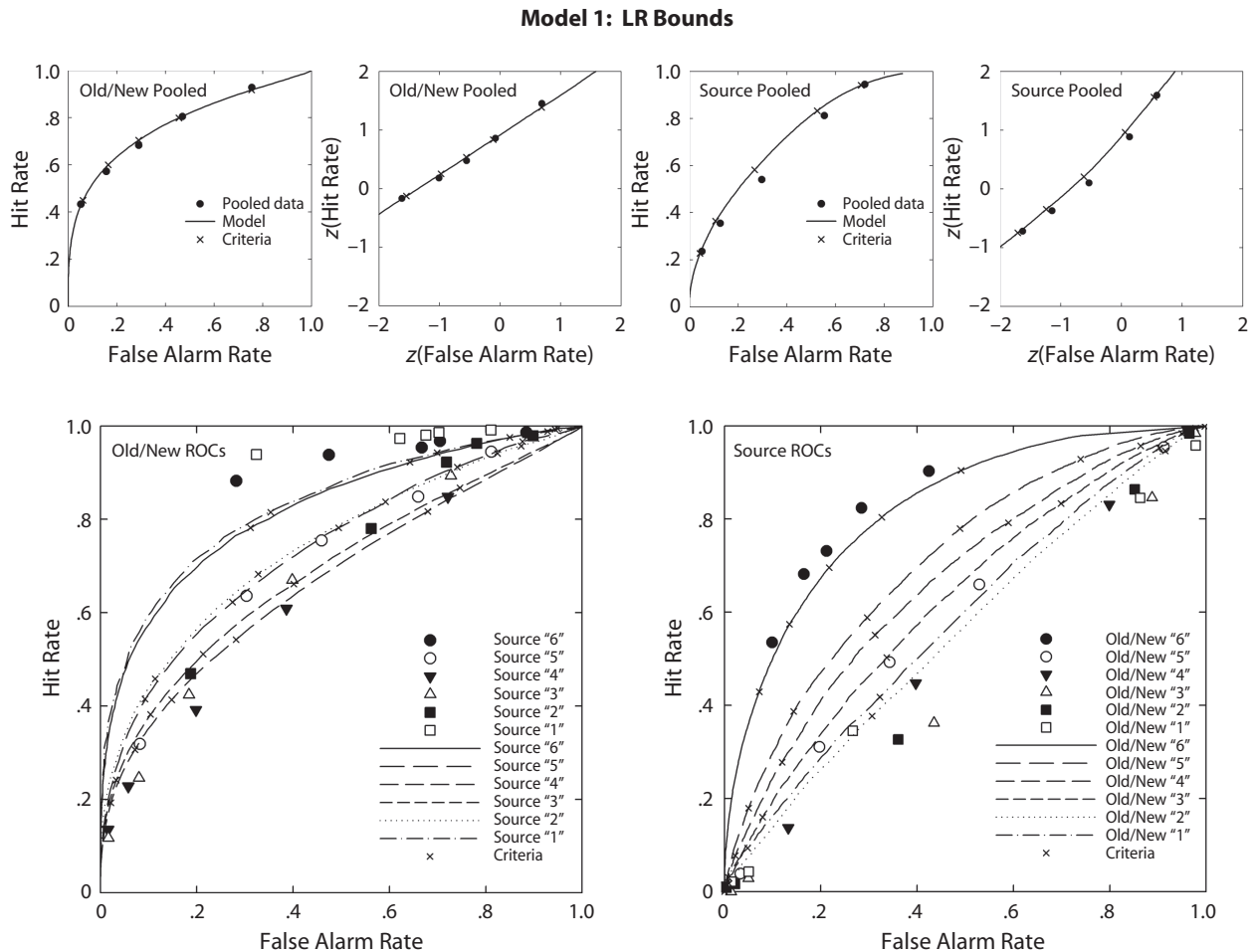
Model 1: LR Bounds

Figure 9. Model 1 fit to the Yonelinas (1999, Experiment 2) ROC data, arranged analogously to Figure 7.

lead to conclusions discrepant from those of the Yonelinas (1999) study. In addition, complete information about the Slotnick et al. (2000) experiments is available online (www.psychonomic.org/archive).

Model 2: Likelihood-Ratio Bounds and Guessing After "New" Responses

The dual-rating task used by Yonelinas (1999) and Slotnick et al. (2000) has a feature that renders Model 1 counterintuitive: No matter what level of old–new confidence is assigned to an item, the subject must provide a confidence judgment concerning the source. The assumptions of continuous distributions and likelihood-based decision boundaries force the prediction that source judgments following "new" decisions were not driven by chance: Subjects must have responded "female" at different rates to items presented in a female or a male voice. A plausible alternative hypothesis is that subjects viewed source judgments following "new" responses as meaningless, and therefore simply guessed rather than consulting decision boundaries.

To implement this idea, we let the probabilities of guessing particular source levels take on arbitrary values, but required that these values be the same for all "new"

responses (ratings 1, 2, and 3 for the Yonelinas, 1999, data and 1–4 for the Slotnick et al., 2000, data[9]). Five guessing parameters were therefore required for the Yonelinas (1999) data set, and six for Slotnick et al. (2000). As Table 2 shows, the guessing assumption improved the fit substantially over Model 1 for all three data sets. RMSEA is reduced by a factor of more than 2, and G^2 by a factor of more than 3, for all data sets. Because Model 2 has more parameters than Model 1, AIC and BIC values were calculated; smaller values of either measure indicate a better fit. As shown in Table 3, Model 2 is superior to Model 1 by either standard for all data sets.

Figure 10 shows the ROCs for the Yonelinas (1999) data, together with fits of the new model. The refined ROCs (especially the old–new curves) fall closer to the points than for Model 1, and the predicted locations of points along each curve are also more accurate. For Model 1, even points at the correct level of sensitivity were far from the predicted values (indicated by ×s on the graphs), but Model 2 gets it right. Apparently, Model 1 had difficulty with this aspect of the data because of its unrealistic assumption that LR bounds were in use in the "new" region.[10]

Table 3
AIC and BIC Values

Model	Yonelinas (1999)		Slotnick et al. (2000, Experiment 2)		Slotnick et al. (2000, Experiment 3)	
	AIC	BIC	AIC	BIC	AIC	BIC
DeCarlo (2003b)	2,456	2,507	1,624	1,687	3,807	3,870
LR bounds	989	1,040	662	725	1,228	1,291
LR + guessing	261	326	382	462	432	512
LR, guessing, and inattention	236	314	310	406	427	523

Note—The values were found from the G^2 values in Table 2 by adding penalties for the number (k) of free parameters: $2k$ for AIC and $k\ln(n)$ for BIC, where n is the number of cells in the data matrix. Calculations based on log likelihoods yielded an identical pattern. AIC, Akaike information criterion; BIC, Bayesian information criterion.

In a recent article relevant to the guessing question, Starns, Hicks, Brown, and Martin (2008) studied source identification of unrecognized items. In three experiments, they found performance to be above chance, provided that the old–new response criterion was conservative (high). Their conservative criteria (obtained by informing subjects that only 25% of the test items had been studied) were nearer to the mean of the target distributions than to that of the lure distribution, and thus correspond to ratings in the "old" category on our scale. For their liberal condition (in which subjects were told that 75% of test items had been studied), source identification was at chance. Old–new criteria in those conditions were nearer to the lure distribution mean, and thus correspond to the "new" ratings on our scale. Thus Starns et al.'s findings support the assumption that the subjects in the experiments we are analyzing, unable to reach above-chance source performance for probes they believed to be "new," fell back on source guessing for those items.

Model 3: Likelihood-Ratio Bounds, Guessing After "New" Responses, and Reduced-Attention Noise

Model 2 does quite well, but there is room for improvement. One possible weakness was noted by DeCarlo (2003b), who speculated that the poor fit of his correlated-distributions model to the Yonelinas (1999) and Slotnick et al. (2000) data sets might lie in "violations of distributional assumptions" (De Carlo, 2003b, p. 295). In particular, he conjectured that his own attentional model (DeCarlo, 2003a), in which the effective target distributions are mixtures of two normal distributions, might improve the fit. Slotnick and Dodson's (2005) "added noise" concept, a variant of the mixture model assumption, allowed them to account for changes in accuracy and in ROC shape with old–new confidence. However, their analyses were still univariate: Even refined ROCs model only one dimension (source rating), factoring out (rather than averaging over) the other dimension (old–new rating). Model 2 accounts well for the pooled ROCs but not for the refined curves, especially the lower ones. Incomplete attention might account for these discrepancies; our next model incorporates the conjecture that a mixture of full and reduced attention underlies the data sets we have been considering.

To the decision space represented in Figure 8, we added a noise distribution (U). When a previously studied item

is tested, its strength is sampled from the appropriate target distribution on some (attended) trials, but from the U distribution on other (reduced-attention) trials. A given source judgment no longer depends simply on the relevant source distribution, but rather on a mixture of that distribution and U. We saw earlier that mixture models of this type degrade source discrimination (DeCarlo, 2003a). For our combined old–new/source model, this degradation is increasingly severe at lower levels of old–new confidence, because those ratings are more likely to arise on unattended (U) trials. There is little degradation for confident "old" judgments.

We therefore modified the model to include a noise distribution, a change that required five additional parameters: a mean and variance on both dimensions, and the fraction of test trials on which the distribution was invoked.[11] (The bivariate correlation of the noise distribution was assumed to be 0.) The resulting parameter estimates are again given in Table 1. For the Yonelinas (1999) data, the best fit implied a noise distribution with mean 0.90 units above and 0.32 units to the right of the lure distribution, suggesting that the added noise interfered more with old–new decisions than with source judgments. The variance parameters (as compared with the lure distribution variance of 1) were $\sigma_x^2 = 0.92$ and $\sigma_y^2 = 1.06$. In Slotnick et al.'s (2000) Experiment 2, the noise means were $\mu_x = 0.0$ and $\mu_y = 1.30$, and the variances were $\sigma_x^2 = 2.45$ and $\sigma_y^2 = 2.26$; again, the noise affected old–new judgments more than source decisions. Experiment 3 showed a similar pattern: $\mu_x = -0.18$, $\mu_y = 1.02$, $\sigma_x^2 = 1.18$, $\sigma_y^2 = 1.12$. The proportions of unattended trials were .46 for the Yonelinas (1999) data set, and .18 and .25 for the two Slotnick et al. (2000) data sets. The lower proportions of inattention estimated from the Slotnick data are consistent with the higher levels of sensitivity obtained in their experiments.

In Figure 11, which shows the fit to the Yonelinas (1999) data, the curvature in the pooled recognition zROC data is now reflected in the model, and the spacing of the refined curves more nearly captures the data than was the case with the previous models. The addition of the noise component improved the fits to all three data sets, as can be seen in Table 2. Because Model 2 is nested within Model 3, the reliability of the improvement can be evaluated: The difference in G^2 values between conditions is approximately distributed like a χ^2 with dfs equal
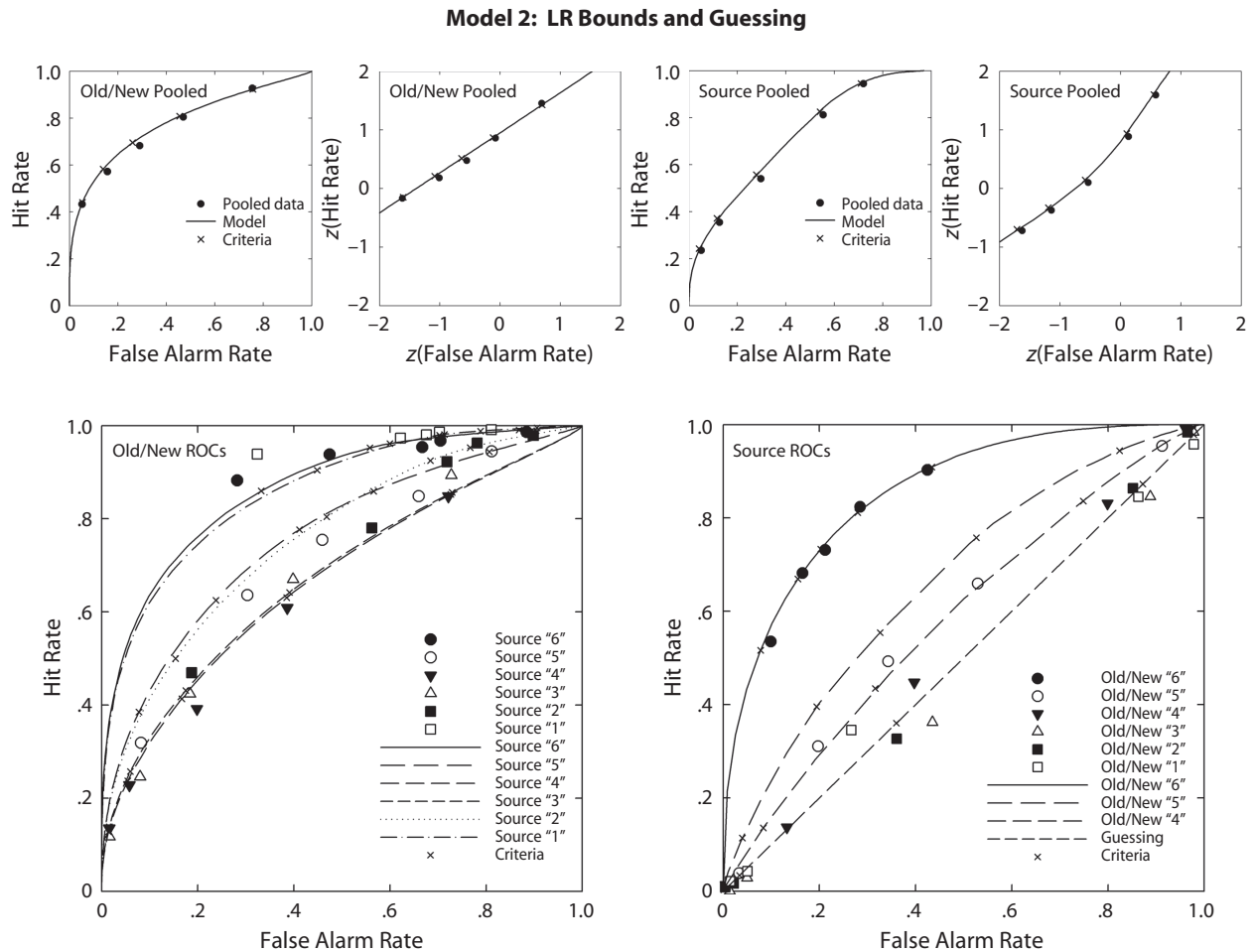
Model 2: LR Bounds and Guessing

Figure 10. Model 2 fit to the Yonelinas (1999, Experiment 2) ROC data, arranged analogously to Figure 7.

to the difference in *df*s between the models (in this case, $\Delta df = 5$). For the three data sets, $\Delta G^2 = 35, 26$, and 19; all of these values are highly reliable ($p < .005$), indicating that the addition of the U distribution has indeed improved our ability to describe the data. Values of RMSEA in the three data sets are also reduced, in all cases by 10% or more. The AIC and BIC statistics (Table 3) support this conclusion for two of the three studies, but for Slotnick et al.'s (2000) Experiment 3, the AIC difference is small, and the BIC actually favors Model 2.

Another recent model in which noise not equivalent to that of the lures has been postulated is the "misattribution" proposal of Dodson, Bawa, and Slotnick (2007). In their (one-dimensional) representation, the distribution resulting from items spoken in a male voice was a mixture of the true distribution and one that was equivalent to that produced by the female voice. One distinction between their model and Model 3 is that the location of the added distribution was postulated by Dodson et al. but is determined by parameter search in our approach. Our modeling located the U distribution an average of only 0.05 units from 0 across the three data sets. This result suggests that the effect of inattention is to reduce sensitivity overall,

rather than to shift some of the mass from the correct distribution to the erroneous one.

DISCUSSION

We have described a set of decision models for source memory experiments that include item recognition judgments. The best of these models is successful at its task. But how do the processes postulated by these models fit into broader theorizing about source judgments? In this final section, we first summarize what we take to be the most important characteristics of the models and briefly consider the implications of these features for further research. We then contrast our approach with two alternatives that have dominated the literature.

Aspects of the Models: (Two-Dimensional) Strength, Likelihood-Ratio Bounds, and Nonoptimal Components

According to the models explored in this article, source judgments are made by comparing the strength of a probe with the distributions resulting from the encoding of items in memory. Two kinds of strength mediate these decisions:
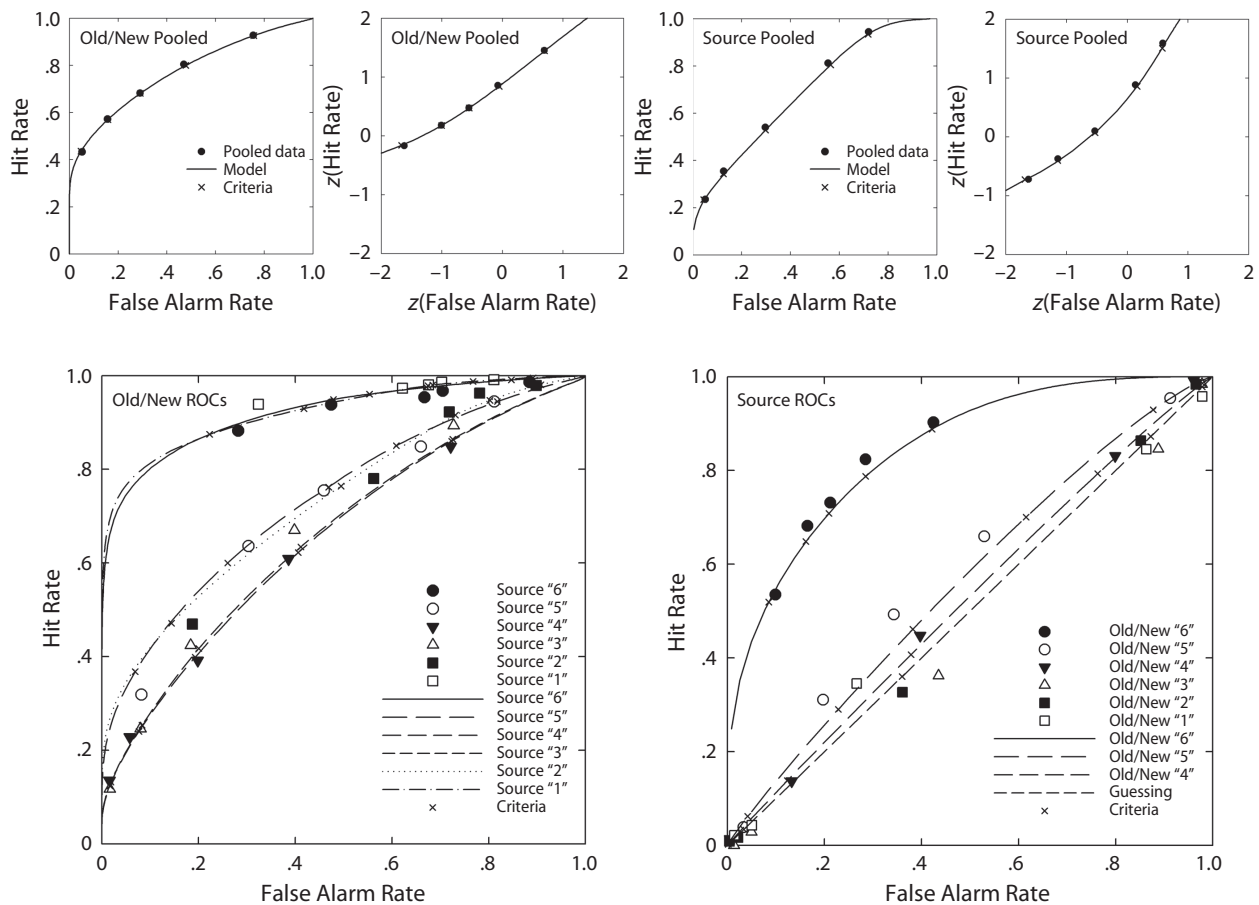
Model 3: LR Bounds, Guessing, and Inattention

Figure 11. Model 3 fit to the Yonelinas (1999, Experiment 2) ROC data, arranged analogously to Figure 7.

the strength of the probe item, which distinguishes old and new test probes, and the relative strengths attributable to the two sources. The source decision and the confidence with which it is made are functions of LR, or the relative probability that the observation resulted from one of the two sources. Although the old–new decision is based on pure strength rather than LR, each judgment depends on the relative strengths that support two hypotheses.

Strength models are a natural starting place in memory research, and Hoffman's (1997) model of source judgments represents the simplest possible model of this type. Because subjects' accuracies in source and item recognition tasks tend to be of similar magnitudes, a strength dimension that underlies source judgments must not be the same as the dimension that supports old–new decisions. This insight led Banks (2000) and DeCarlo (2003b) to propose the two-dimensional representations that were the starting point for our enterprise. Because these models were explicit about representation and decision rules, their failure to provide good fits was informative, and our models have built in natural ways on this foundation.

The greatest improvement in our series of models came from replacing linear bounds with bounds based

instead on LR. Such bounds may seem more complicated than the linear ones in DeCarlo's (2003b) model, but they are "simple" in the sense of being optimal (Green & Swets, 1966, ch. 1, give several criteria for optimality of binary decisions that are satisfied by LR). An unexpected payoff of employing LR bounds was the finding that the resulting source zROCs were concave upward, an empirical result that had been difficult to account for in signal detection terms. In fact, this finding has repeatedly been cited as evidence for a threshold component in source recognition (see, e.g., Yonelinas & Parks, 2007). Yet, the fits provided by Model 1 show that a pure signal detection model, albeit in two dimensions, can produce the effect.

Adding two nonoptimal processes, in Models 2 and 3, led to increasingly better descriptions of the data (except that Model 3 provided no incremental improvement for the Slotnick et al., 2000, Experiment 3 data). These two processes have somewhat different status. The assumption that subjects guess the source of an item when they have just denied recognizing it may not be technically optimal, but surely it is cognitively defensible. Incomplete attention (Model 3), on the other hand, can only be counted as

flawed encoding. Nonetheless, support for mixture models of this type has been found repeatedly in source memory experiments (DeCarlo, 2003a; Hilford et al., 2002).

This set of models allows us to ask questions about the effects of payoffs, presentation probabilities, and other experimental manipulations designed to influence "response bias." Such biasing variables can be applied to either the old–new or the source judgment. Because all bounds are expected to shift to higher or lower values of LR, the form of the various ROCs should be unaffected. More interestingly, what changes would be expected according to Model 3? For example, if two thirds of the items were from source A, the distribution due to inattention (U) might shift along the source axis.[12]

Alternative Approaches

One-dimensional signal detection analyses. The two-dimensional framework can be contrasted, to its advantage, with the popular one-dimensional approach to modeling source memory. In the introduction, we noted the controversy over the "shape" of the source ROC and its theoretical basis. As different shapes have been discovered, corresponding models have been proposed to account for them. Aspects of shape have been considered to be diagnostic of the models; for example, upwardly curved zROCs were viewed as support for threshold models. Until source judgments and the resulting ROCs were studied together with old–new judgments, models that reflected different aspects of a larger phenomenon appeared to be in conflict.

This search for signature predictions of particular models has two other shortcomings. First, focusing attention on one particular aspect of the data necessarily shortchanges other elements, which are implicitly understood to be of lesser importance. Second, because such predictions refer to only part of the data, they encourage the use of restricted designs (e.g., studies of source memory that do not include old–new judgments). Restricted designs run the risk of providing a restricted view that supports conclusions of limited generality.

A particular strategy that has been used in conjunction with one-dimensional SDT analyses illustrates another advantage of our approach. Dual-process, signal detection, and mixture SDT models all predict ROCs and zROCs with particular types of curvature, and the presence of such curvature has been measured by the magnitude of quadratic terms in polynomial fits (Parks & Yonelinas, 2007). But none of these models predict curvature of a quadratic form. Our models provide good descriptions of the multiple types of curvature that are observed, and these models can be used to infer the processes that influence ROC shape.

A final consequence of limited modeling is that large frameworks such as SDT may be rejected on the basis of very limited subcases, even though a fuller consideration of the data strongly supports that same framework. The introduction mentioned the many articles that have contrasted threshold and signal detection views, but in all of these, only one-dimensional SDT was considered. We hope to have shown that two-dimensional detection-theoretic models capable of capturing item and source recognition phenomena are natural extensions of fundamental SDT principles (using such elements as continuous representation and LR decision rules) and cannot be evaluated by testing the predictions of restricted models.

Models with explicit recollection processes. In the literature, source monitoring is typically conceived of as depending on qualitative stimulus characteristics that are "diagnostic" of a particular source (Johnson et al., 1993). For example, in reality monitoring, a high degree of perceptual detail evoked by a probe may serve to identify that item as having been seen rather than imagined. In one common elaboration of this idea, source judgments are driven by a recollective process (Yonelinas, 1999), and the most striking aspect of the present models for many readers is likely to be the complete absence of such a construct.

In their recent review of ROCs in recognition memory, Yonelinas and Parks (2007) concluded that "single-component models of recognition memory are inadequate" (p. 829). The success of the present models of source identification might be taken to support this view, depending on whether the second dimension in Figure 8 is taken to be a second "component." For many researchers, however, the required renovation of simple models is the inclusion of a separate "recollection" process. A view popular with both dual-process and SMF theorists is that the ability to assign a source to a memory is an indication that recollection, rather than mere familiarity, accounts for recognition (Hicks, Marsh, & Ritschel, 2002; Yonelinas, 1999). Indeed, accurate source judgments are sometimes taken as evidence for the use of recollection (Chan & McDermott, 2007; Yonelinas, 2002).

Without specifying the characteristics of recollection, this view cannot be distinguished from our signal detection models, because the dimension that distinguishes the sources could simply be renamed *recollection*. In fact, Wixted (2007; Wixted & Stretch, 2004) has argued that recollection may combine with familiarity to generate a single item-strength decision variable, so that each old–new judgment depends on a sum of the two. Rotello, Macmillan, and Reeder (2004) proposed a similar decision rule. What makes the dual-process view of recollection special is the assumption that it is a high-threshold process.

Wixted (2007) recently argued that the pattern of refined-ROC shapes is strongly inconsistent with the threshold-generated predictions of dual-process theory, and specifically with linear source ROCs (Yonelinas, 1999). Because the dual-process model asserts that recollection generates the highest-confidence responses, it predicts that the shape of refined ROCs should change from linear at a high level of old–new confidence to SDT-like at low confidence, exactly the opposite of what is observed (Slotnick & Dodson, 2005). In reply, Parks and Yonelinas (2007) argued that at high levels of confidence, the familiarity process would exhibit high sensitivity, accounting for the observed pattern. This effect is not explicitly predicted by the dual-process model, and is contrary to the spirit of Yonelinas's (1999) initial investigation of source ROCs.

Furthermore, and finally, consider how well the data are described by the models presented here. The contribution that might be made by adding threshold components is limited to the small fraction of the variance not already explained. We conclude that there is nothing in the classic data sets we have modeled that can be taken as support for a separate, threshold-based recollection process. Signal detection processes that depend on two aspects of memory strength are enough.

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: Akademiai Kiado.

ASHBY, F. G., & TOWNSEND, J. T. (1986). Varieties of perceptual independence. *Psychological Review*, **93**, 154-179.

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

BINK, M. L., MARSH, R. L., & HICKS, J. L. (1999). An alternative conceptualization to memory "strength" in reality monitoring. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **25**, 804-809.

BROWNE, M. W., & CUDECK, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136-162). Newbury Park, CA: Sage.

CHAN, J. C. K., & MCDERMOTT, K. B. (2007). The testing effect in recognition memory: A dual process account. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **33**, 431-437.

COHEN, A. L., ROTELLO, C. M., & MACMILLAN, N. A. (2008). Evaluating models of remember–know judgments: Complexity, mimicry, and discriminability. *Psychonomic Bulletin & Review*, **15**, 906-926.

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, & 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.

DODSON, C. S., BAWA, S., & SLOTNICK, S. D. (2007). Aging, source memory, and misrecollections. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **33**, 169-181.

GLANZER, M., HILFORD, A., & KIM, K. (2004). Six regularities of source recognition. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **30**, 1176-1195.

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

HICKS, J. L., MARSH, R. L., & RITSCHEL, L. (2002). The role of recollection and partial information in source monitoring. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **28**, 503-508.

HILFORD, A., GLANZER, M., KIM, K., & DECARLO, L. T. (2002). Regularities of source recognition: ROC analysis. *Journal of Experimental Psychology: General*, **131**, 494-510.

HOFFMAN, H. G. (1997). Role of memory strength in reality monitoring decisions: Evidence from source attribution biases. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **23**, 371-383.

JOHNSON, M. K., HASHTROUDI, S., & LINDSAY, D. S. (1993). Source monitoring. *Psychological Bulletin*, **114**, 3-28.

JOHNSON, M. K., & RAYE, C. L. (1981). Reality monitoring. *Psychological Review*, **88**, 67-85.

KINCHLA, R. A. (1994). Comments on Batchelder and Riefer's multinomial model for source monitoring. *Psychological Review*, **101**, 166-171.

PARKS, C. M., & YONELINAS, A. P. (2007). Moving beyond pure signal-detection models: Comment on Wixted (2007). *Psychological Review*, **114**, 188-201.

PITT, M. A., MYUNG, I. J., & ZHANG, S. (2002). Toward a method of selecting among computational models of cognition. *Psychological Review*, **109**, 472-491.

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, & Cognition*, **27**, 1110-1115.

ROTELLO, C. M., MACMILLAN, N. A., & REEDER, J. A. (2004). Sum–difference theory of remembering and knowing: A two-dimensional signal-detection model. *Psychological Review*, **111**, 588-616.

SCHWARZ, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, **6**, 461-464.

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, & Cognition*, **26**, 1499-1517.

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.

TANNER, W. P., JR. (1956). Theory of recognition. *Journal of the Acoustical Society of America*, **28**, 882-888.

WIXTED, J. T. (2007). Dual-process theory and signal-detection theory of recognition memory. *Psychological Review*, **114**, 152-176.

WIXTED, J. T., & STRETCH, V. (2004). In defense of the signal detection interpretation of remember/know judgments. *Psychonomic Bulletin & Review*, **11**, 616-641.

WONG, M. (2002). *Is source memory continuous or discrete? An ROC analysis.* Unpublished master's thesis, University of Massachusetts, Amherst.

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, & Cognition*, **25**, 1415-1434.

YONELINAS, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory & 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.

NOTES

1. Examination of Figure 2B reveals a systematic discrepancy between the data and the Gaussian model, so in spite of the presence of "curvature," the Qin et al. (2001) data support that model only weakly.

2. Dodson, Bawa, and Slotnick (2007) proposed an alternative mixture model in which inattention is replaced by "misattribution." We comment on this proposal later.

3. For simplicity, Figure 6 shows only two criteria of each type.

4. DeCarlo (2003b) allowed the bivariate correlation of the lure distribution to vary, but in our fits the obtained value was always near 0, so we fixed it at that value. Also, DeCarlo (2003b) used four parameters to specify the x and y locations of the target distributions, but we reduced this number to three (without loss of generality) by forcing the x locations of the target distributions to be the same magnitude but opposite in sign.

5. The locations of the decision criteria are omitted.

6. The complete data matrices for Experiment 2 of Yonelinas (1999) and Experiments 2 and 3 of Slotnick et al. (2000) were published by Slotnick and Dodson (2005).

7. For consistency, the old–new decision bounds should also be based on LR, but we did not fit such bounds. We made this choice in part because the difference between linear and LR old–new bounds is not as great as the difference for source bounds, but our primary motivation was that finding the best-fitting values of LR for both sets of bounds simultaneously proved intractable.

8. Like the DeCarlo (2003b) model, Model 1 has 19 parameters for the Yonelinas (1999) data and 21 for Slotnick et al. (2000). Criterion values account for 10 (Yonelinas) or 12 (Slotnick) of these. Although this seems like a large number, almost as many parameters are required by the conventional one-dimensional approach (if both old–new judgments and source recognition are measured), and that strategy is unable to model interactions between the two subtasks.

9. We also considered treating just responses of "1"–"3" in the Slotnick data in this way. However, the data appear more consistent with guessing following "4" responses as well, and modifying the model to assume guessing only after "1"–"3" produced a poorer fit.

10. An alternative to the guessing hypothesis, pointed out by a reviewer, is that the underlying distributions may be something other than bivariate normal. The below-chance predictions of Model 1 for refined source ROCs arise because two normal distributions with correlation terms of opposite sign necessarily cross over, but the actual evidence distributions need not follow this mathematical prescription. Since we have no intuitions to guide the choice of alternative distribution shapes, we do not pursue this possibility.

11. It might seem that such a distribution would necessarily have the same properties as the lure distribution. We did compare a model of this type with a version of Model 3 (in both cases, without the guess-ing assumption) and found that leaving the means and variances as free parameters produced a reliably better fit. Furthermore, the estimated parameters of this distribution have a natural interpretation, reducing the risk that we are simply accounting for random variation in the data.

12. We thank an anonymous reviewer for pointing out this avenue of investigation.

ARCHIVED MATERIALS

The following materials associated with this article may be accessed through the Psychonomic Society's Norms, Stimuli, and Data archive, www.psychonomic.org/archive.

To access these files, search the archive for this article using the journal name (*Psychonomic Bulletin & Review*), the first author's name (Hautus), and the publication year (2008).

FILE: Hautus-PB&R-2008.zip.

DESCRIPTION: The compressed archive file contains one file, ROC_fits_to_Slotnick_et_al_2000.pdf, a PDF document that displays observed and predicted ROCs for the data from Slotnick et al. (2000). The format is the same as in Figures 7 and 9–11 in this article.

AUTHOR'S E-MAIL ADDRESS: m.hautus@auckland.ac.nz.

AUTHOR'S WEB SITE: www.psych.auckland.ac.nz/people/Hautus/Hautus.htm.