Journal of Experimental Psychology:
Learning, Memory, and Cognition
2000, Vol. 26, No. 6, 1499–1517

0278-7393/00/$5.00 DOI: 10.1037//0278-7393.26.6.1499

An Analysis of Signal Detection and Threshold Models of Source Memory

Scott D. Slotnick
Johns Hopkins University

Stanley A. Klein
University of California, Berkeley

Chad S. Dodson
Harvard University

Arthur P. Shimamura
University of California, Berkeley

The authors analyzed source memory performance with an unequal-variance signal detection theory model and compared the findings with extant threshold (multinomial and dual-process) models. In 3 experiments, receiver operating characteristic (ROC) analyses of source discrimination revealed curvilinear functions, supporting the relative superiority of a continuous signal detection model when compared with a threshold model. This result has implications for both multinomial and dual-process models, both of which assume linear ROCs in their description of source memory performance.

Source memory refers to memory for the context in which information was acquired (Johnson, Hashtroudi, & Lindsay, 1993). For example, memory for the person with whom one is conversing or the place where one is conversing can be interpreted as source memory. In psychological experiments, source memory is typically assessed by asking participants to determine the origin of previously presented information, such as whether the information was presented verbally or visually, presented by a male or a female voice, or presented in one spatial location or another. As these examples imply, source memory depends on memory for autobiographical or episodic information. Various cognitive and neuropsychological findings have suggested that, to some degree, memory for source can be dissociated from item memory (see Dodson & Shimamura, 2000; Johnson, Kounios, & Reeder, 1994; Schacter, Harbluk, & McLachlan, 1984; Shimamura & Squire, 1987; Zaragoza & Lane, 1994). Indeed, various models of memory suggest a distinction that is related to differences between item and source memory (e.g., Gardiner, 1988; Hirst, 1982; Jacoby, 1991; Johnson et al., 1993; Mayes, Meudell, & Pickering, 1985; Tulving, 1972).

Johnson et al. (1993) developed a useful framework for the analysis of source memory. In this "source monitoring" framework, the degree to which individuals identify the source of a memory depends, in part, on the kind of information that is acquired and remembered. That is, one can remember various aspects of a learning episode, such as perceptual information, spatial information, semantic detail, affective information, and the cognitive operations invoked during learning. Some of these aspects, such as perceptual and spatial (i.e., contextual) information, may be particularly important for making a correct source attribution. For example, remembering the particular quality of a speaker's voice may facilitate identifying which speaker presented some information. The source monitoring framework provides a useful characterization of both the features of episodic memory that are important for the establishment of source memory and the decision processes that are involved in the retrieval of these memories.

In another line of research, formal models have been developed for evaluating the different processes associated with source monitoring. Batchelder and Riefer (1990; Batchelder, Riefer, & Hu, 1994) have developed a multinomial modeling approach that can be used to derive parameters associated with item memory, memory for source, and guessing biases. Various modifications of the original Batchelder–Riefer model have been applied successfully to address issues of source memory and related phenomenon, such as recollective processes and response bias (see Bayen, Murnane, & Erdfelder, 1996; Buchner, Erdfelder, Steffens, & Martensen, 1997; Dodson, Holland, & Shimamura, 1998; Dodson & Shimamura, 2000; Erdfelder & Buchner, 1998). Recently Yonelinas (1999) used the threshold recollective component of the dual-process model (Jacoby, 1991) to describe source memory performance. Both the Batchelder–Riefer model and dual-process model evaluate memory for source in terms of a three-state or two-high threshold model in which participants either (1) remember that the information came from one source (Source A), (2) remember that the information came from another source (Source B), or (3) do not remember the source and guess.

Threshold models have a rich history in cognitive research (see Banks, 1970; Snodgrass & Corwin, 1988). Thus, it is reasonable that source memory has been considered in terms of such models. An important advantage of this modeling approach is that it is possible to dissociate the contributions of item memory (detection) and memory for source (identification). Moreover, it is possible to

Scott D. Slotnick, Department of Psychology, Johns Hopkins University; Stanley A. Klein, Department of Vision Science, University of California, Berkeley; Chad S. Dodson, Department of Psychology, Harvard University; Arthur P. Shimamura, Department of Psychology, University of California, Berkeley.

This research was supported by National Institutes of Health Grants MH48757, NS17778, and EY04476.

Correspondence concerning this article should be addressed to Scott D. Slotnick, Department of Psychology, Johns Hopkins University, 3400 North Charles Street, Baltimore, Maryland 21218. Electronic mail may be sent to slotnick@jhu.edu.

study the manner in which guessing biases influence memory performance (see Riefer, Hu, & Batchelder, 1994).

The threshold approach is not the only way to assess memory performance. Signal detection theory offers an alternative approach that does not assume that individuals adopt a discrete "state" of knowledge. In signal detection models, the distributions of old and new items are assumed to be gaussian and extend along a single dimension. As is well known, signal detection approaches have been applied frequently to analyses of perceptual phenomena (Egan, 1958; Green & Swets, 1966; for a general review of signal detection theory, see Macmillan & Creelman, 1991). Moreover, this approach has also been used to describe various memory phenomena (Atkinson & Juola, 1974; Banks, 1970; Banks, 2000; Donaldson, 1996; Ratcliff, McKoon, & Tindall, 1994; Ratcliff, Sheu, & Gronlund, 1992; Snodgrass & Corwin, 1988; Yonelinas, 1994).

The present investigation assessed the appropriateness of signal detection theory to source memory. That is, we examined whether source identification can be viewed in terms of a continuous process in the same manner by which item detection has been construed. In addition, we compared a signal detection approach to the threshold approach. In this way, we could evaluate the advantages and disadvantages of both approaches.

Threshold Models

The Batchelder–Riefer model of source monitoring has been described in detail elsewhere (see Batchelder & Riefer, 1990; Bayen et al., 1996; Dodson, Prinzmetal, & Shimamura, 1998; Dodson, Holland, et al., 1998). Thus, we only present a summary of the model's basic tenets. In a typical source memory experiment, participants acquire information from two sources, identified as Source A and Source B. Source A and Source B items could be words presented by a male and a female voice, respectively. At test, participants are presented with a mixed list of Source A, Source B, and new words. For each test item, they are asked to make a three-choice source recognition judgment in which they must determine whether a test item came from Source A, Source B, or was not presented at study (i.e., a new item). As shown in Table 1, the data set from such a source test can be summarized in a 3 × 3 confusion matrix. The rows in Table 1 correspond with the three types of items that were presented during test (Source A, Source B, new items), and the columns correspond with the three types of responses for a particular test item.

Table 1
Data Set From Multinomial Source Memory Approach

	Participant response		
Given	"Source A"	"Source B"	"New"
Source A item	**p("A"|A)**	p("B"|A)	p("New"|A)
Source B item	p("A"|B)	**p("B"|B)**	p("New"|B)
New item	p("A"|New)	p("B"|New)	**p("New"|New)**

Note. Bold items represent correct responses. "Source A" = Responding that an item came from Source A; "Source B" = Responding that an item came from Source B; "New" = Responding that an item was new.

The Batchelder–Riefer multinomial approach uses the confusion matrix to derive parameters associated with item detection, source identification, and guessing biases. The parameter space is defined by a decision tree structure, such as that outlined in Figure 1, which displays the memory states that are associated with responding to studied items (i.e., items from Sources A and B) and new items. As seen in Figure 1, there are two memory parameters: (1) item detection (D) refers to the memorial information that allows studied words to be distinguished from new words on the test and (2) source identification (d) refers to the memorial information that identifies the source of studied words. Various guessing processes influence performance when participants fail to remember item or source information. The tree structure in Figure 1 illustrates the different contributions of memory and guessing parameters for each response category in Table 1. Most important, d_a refers to the probability of recollecting an item from Source A and d_b refers to the probability of recollecting an item from Source B.

Bayen et al. (1996) proposed a modified version of the model developed by Batchelder and Riefer. In the original Batchelder–Riefer model, source identification was viewed as a two-high threshold process, whereas item detection was viewed as a one-high threshold phenomenon. In the Bayen et al. (1996) modification, source identification and item detection are both viewed as two-high threshold processes. Specifically, the revised model adds a parameter for the detection of new items (D_n). The tree structure for new items shown in Figure 1 includes the D_n parameter and thus represents the Bayen et al. (1996) modification to the original Batchelder–Riefer model.

Because a full description of the parameters associated with the dual-process model is covered elsewhere (Yonelinas, 1999), we only discuss the model parameters as they relate to source memory performance. In a source memory paradigm that consists of two sources of approximately equal familiarity (e.g., words spoken by either a male or a female in random order), familiarity cannot be used for source identification. Under these conditions, the dual-process model uses only two recollection parameters, R_t and R_l, where R_t refers to the probability of recollecting an old item and R_l refers to the probability of identifying a new item. Yonelinas (1999) used this reduced model, where the familiarity component is set to zero, in three of four experiments where the process of familiarity was not expected to influence source identification. Like the Batchelder–Riefer model, the reduced dual-process model assumes a two-high threshold process underlies source identification. Therefore, when sources are of similar familiarity, fitting the two-high threshold model to source identification data is the same as fitting both the Batchelder–Riefer model and the dual-process model.

Continuous Signal Detection Models and the Analysis of Receiver Operating Characteristics

Threshold models of memory are attractive because of their simplicity (Bernbach, 1967). Moreover, in many instances they are sufficient for analyses of empirical data, particularly those used to suggest qualitative changes, such as the identification of functional dissociations between cognitive or neuropsychological variables. Yet, continuous signal detection models have certain advantages, especially in modeling finer-grain, quantitative aspects of memory

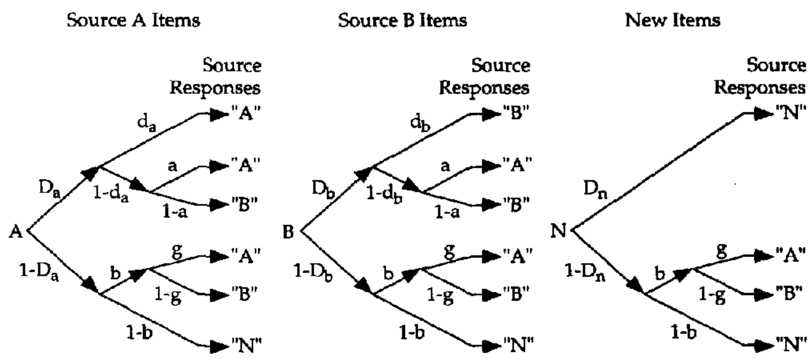

Figure 1. Tree diagrams for the two-high threshold multinomial model, with separate trees for Source A items, Source B items, and New items. D_i = probability of detecting a Source i item as old; d_i = probability of correctly identifying the item as originating from Source i (i refers to Source A or Source B); D_n = probability of detecting a New item as new; a = probability of guessing that a detected item is from Source A; b = probability of guessing an item is old; and g = probability of guessing that an undetected item is from Source A.

performance (see Drake & Hannay, 1992; Ratcliff et al., 1992; Yonelinas, 1994).

In standard applications of signal detection theory to memory performance, individuals are asked to make confidence ratings as to whether a particular test word was an "old" or "new" item. For example, old–new judgments may be obtained by asking individuals to rate on a 7-point scale their confidence that a given test item was old or new (7 = *very sure it was "old"*; 1 = *very sure it was "new"*). From these ratings, a receiver operating characteristic (ROC) can be obtained that characterizes the relationship between hits and false alarms across various levels of confidence. In addition, d' can be calculated, which provides a measure of memory strength in terms of the separability of the gaussian distributions of old (signal) and new (noise) items. Finally, a criterion parameter (C) can be calculated that serves as a bias measure indicating the degree to which both old and new items are endorsed as "old." This measure of confidence bias has been shown to be independent of d' and thus has been recommended for use in place of beta (Snodgrass & Corwin, 1988). C has a value of zero with no bias, is negative for lax bias, and is positive for strict bias.

Signal detection theory has also been used to analyze the underlying variability of old and new items (Ratcliff et al., 1992, 1994). Often, it is assumed that the standard deviation of the old and new distributions are equal to each other and that ROCs are symmetrical along the diagonal. However, it is possible to evaluate the degree to which the variability differs between the old and new distributions. Specifically, a measure of the ratio of the standard deviations of new (σ_n) and old (σ_o) distributions can be obtained by taking the z transform of the hit and false alarm rates for each confidence rating (see Green & Swets, 1966; Macmillan & Creelman, 1991; Ratcliff et al., 1992). If the distributions are assumed to be gaussian, the slope of the z-transformed ROC (zROC) represents a measure of the ratio between the new and old standard deviations (σ_n/σ_o). In practice, this ratio is typically below 1, because the variability of old items is generally greater than the variability of new items. Ratcliff and colleagues (Ratcliff et al., 1992, 1994) have shown that this ratio hovers about .8 across various manipulations of memory strength. This finding poses difficulty for a variety of global memory models, which often predict changes in the variability of old and new distributions as a function of memory strength (Gillund & Shiffrin, 1984; Hintzman, 1986) or predict equal variability of old and new distributions (Murdock, 1982). Recently, Glanzer et al. (1999b) have shown the standard deviation ratio is not constant.

When the signal detection approach is applied to source memory performance, it is assumed that memory strength for source is a continuous, unidimensional variable. Source memory strength can be construed as the degree to which the distribution of Source A and Source B items can be discriminated from each other. In our application of signal detection methodology to source memory performance, participants make three confidence ratings for each test item—one rating for item familiarity (old–new recognition) and two ratings for source memory. For one source rating, individuals rate their confidence that a test item came from Source A (7 = *very sure it came from Source A;* 1 = *not sure it came from Source A*). In the second source rating, individuals rate their confidence that a test item came from Source B (7 = *very sure it came from Source B;* 1 = *not sure it came from Source B*).

It is possible to represent the two source ratings as a single dimension such that one endpoint represents the highest confidence that an item came from Source A and the other endpoint represents the highest confidence that an item came from Source B. For example, the highest rating for an item coming from Source A would be those items given a 7 rating on the "Source A" judgment and a 1 rating on the "Source B" judgment. The second highest ranking for items judged as coming from Source A would be the combined set of items given a 7 rating on the "Source A" judgment and a 2 rating of the "Source B" judgment plus those items given a 6 rating on the "Source A" judgment and a 1 rating on the "Source B" judgment. Figure 2 illustrates how the two source judgments for a particular item can be converted into a single 13-point scale. Items with equivocal or no source memory discrimination are those items that are rated equally on both source ratings. Such items are scored at the midpoint of the scale (i.e., a rating of 7 on the 13-point scale).

Source A Items

Raw "Source B" Ratings

Raw "Source A" Ratings	1	2	3	4	5	6	7
7	1	2	3	4	5	6	7
6	2	3	4	5	6	7	8
5	3	4	5	6	7	8	9
4	4	5	6	7	8	9	10
3	5	6	7	8	9	10	11
2	6	7	8	9	10	11	12
1	7	8	9	10	11	12	13

Sum diagonals

↓

Source A Items

← Judge "Source A" Judge "Source B" →

1	2	3	4	5	6	7	8	9	10	11	12	13	Σ

Repeat for each "Old/New" rating

↓

Source A Items

← Judge "Source A" Judge "Source B" →

"O/N"	1	2	3	4	5	6	7	8	9	10	11	12	13	Σ
7														
6														
5														
4														
3														
2														
1														
Σ														

Repeat for Source B and New Items

Figure 2. The top 7 × 7 matrix refers to source judgments to Source A items for a single "old–new" ("O/N") confidence rating (from 1–7). This matrix can be converted into the middle one-dimensional matrix by summing all the bins containing the same number. The middle matrix is a single row in the bottom matrix, where the each row corresponds to each "old–new" confidence rating.

ROC curves for source memory can be drawn from the 13-point rating scale. For example, the proportion of Source A items given a rating of 1 on the 13-point scale (collapsed across "old–new" judgments) is plotted against the proportion of Source B items given that same rating. Cumulative proportions are calculated by including subsequent rating bins and obtaining proportions for both Source A and Source B items. Thus, in the view of standard signal detection methods, the ROC derived from source judgments represents the degree to which individuals can discriminate between Source A items and Source B items.

The analysis of ROCs for source identification permits a test of the underlying structure of source information. To the extent that memory for source is represented as a continuous gaussian distribution of source knowledge—ranging from strong memory for a Source A item to no memory for that item, or ranging from strong memory for a Source B item to no memory for that item—ROCs should conform to standard curvilinear functions such as those obtained in analyses of item detection. To the extent that source identification represents a discrete, two-high threshold process in which individuals are in one of three states—know the item is from Source A, know the item is from Source B, or have no knowledge of source—then ROCs should be linear. Figure 3 illustrates representative ROCs for the two contrasting models of memory for source.

In our experiments, source memory was assessed by asking participants to judge aspects of a learning episode in which words were presented. In three experiments, words were presented by a male voice or a female voice. By obtaining "old–new," "Source A," and "Source B" ratings, we analyzed ROCs to determine whether source memory should be viewed as a two-high threshold process or whether it is better viewed as a continuous process.

Experiment 1

The primary purpose of this study was to assess source memory by applying signal detection theory. Thus, we obtained both old/new and source judgments at various levels of confidence and used these data to obtain ROCs for item memory and source identification. Source memory was manipulated by presenting words in either a male voice or a female voice.

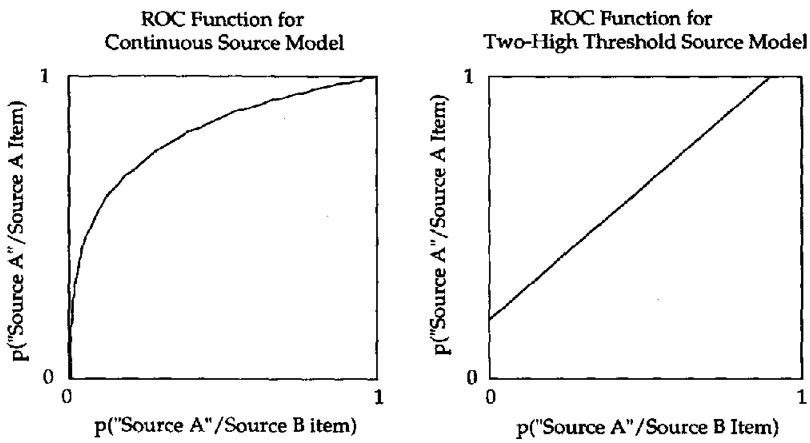

Figure 3. The figure on the left illustrates a curvilinear receiver operating characteristic (ROC) for source identification that is generated from the unequal variance signal detection theory model. The figure on the right illustrates a linear ROC for source identification that is generated from the two-high threshold model.

Method

Participants. Twenty-four undergraduates at the University of California, Berkeley, were each paid $8 for their participation.

Materials. The stimulus set consisted of 120 five-letter nouns. From this set, three 40-word lists were produced, each with an average word frequency of 82.5 (Kučera & Francis, 1967). For each word set, the words were recorded once by a male voice and once by a female voice. The lists were used to construct the three stimulus types—male words, female words, or new words. The words were presented both auditorily and visually. Across participants, all lists were used equally often in the male, female, and new conditions.

Study phase. During the study phase, participants were presented with 80 words, half by a male voice and half by a female voice. They were not informed that there would be a subsequent test of memory, but instead were asked to rate each word according to the difficulty of covertly reproducing or imagining the quality of the voice. Words spoken by the male and female voices were presented in a random order, with the restriction that no more than four of one type were presented in succession. To diminish primacy and recency effects, six additional buffer words were added at the beginning and the end of the lists.

Test phase. All 120 words were presented visually in a random order with the constraint that no more than four of one type were presented in succession. For each word, participants made three judgments on 7-point confidence scales (1 = *not confident;* 7 = *very confident*): (1) they indicated their confidence that the word had been spoken by a male, (2) they indicated their confidence that the word had been spoken by a female, and (3) they indicated their confidence that the word was new. They were also instructed neither to use only ratings of 1 and 7 nor to try to distribute their ratings evenly. The experimenter suggested to participants that they use any rating that corresponded to the strength of their memory.

Results and Discussion

Continuous model parameter estimation. Table 2 displays for each item type (male, female, new) the distribution of ratings for both "old–new" judgments and for source identification judgments. Responses other than the appropriate integer ratings of 1–7 were discarded, resulting in totals less than 960 for each item type. To assess old–new recognition memory, ROCs and zROCs were plotted separately for male versus new items and for female versus new items (see Figure 4). When plotting the item memory ROC, data are summed over the source ratings converting the two-dimensional response matrix into a one-dimensional response matrix. After this, ROC curves are generated by plotting cumulative sums of item and new data against one another. An alternative approach to plotting the ROC operates on the two-dimensional matrices without collapsing over the source ratings (Klein, 1985; Slotnick, 1996). This technique results in slightly higher d' estimates. Best-fit continuous models and two-high threshold models were fit to each ROC.

The ROC curves were fit by chi-square minimization using the Marquardt least squares algorithm (Press et al., 1988). The chi-square function to be minimized is given by:

$$\chi^2 = \Sigma_i(O_i - E_i)^2/\sigma_i^2. \tag{1}$$

The observed data (O_i) is the actual ROC hit rate value, the expected value (E_i) is the hit rate from the model prediction, and the standard deviation is given by binomial statistics:

$$\sigma_i^2 = E_i(1 - E_i)/N \tag{2a}$$

or

$$\sigma_i^2 = O_i(1 - O_i)/N, \tag{2b}$$

where N is the total number in each rating bin. The results using Equations 2a or 2b are almost identical for all the plots evaluated. Equation 2b was used because Equation 2a is more sensitive to the choice of initial condition for the least squares search.

The two-high threshold model assumes that the hit rate probability, E_i, is a linear function of the false alarm rate, F_i:

$$E_i = aF_i + b. \tag{3}$$

The signal detection model assumes that the z score of the hit rate, $z(E_i)$, is a linear function of the z-score of the false alarm rate, $z(F_i)$:

$$z(E_i) = az(F_i) + b. \tag{4}$$

Table 2
Item Detection and Source Ratings (Experiment 1)

Male Voice

← Judge "Female" Judge "Male" →

"O/N"	1	2	3	4	5	6	7	8	9	10	11	12	13	Σ
1	24	20	29	43	47	30	57	40	48	50	66	31	128	**613**
2	0	1	13	4	11	17	23	19	13	6	13	1	1	**122**
3	1	0	1	0	4	4	15	6	8	1	1	1	0	**42**
4	0	0	1	0	2	6	27	10	2	4	1	1	0	**54**
5	0	0	1	0	1	1	14	3	3	1	1	1	0	**26**
6	0	0	0	0	1	1	26	3	3	1	1	1	0	**37**
7	1	0	0	0	1	1	50	0	0	0	1	0	0	**54**
Σ	**26**	**21**	**45**	**47**	**67**	**60**	**212**	**81**	**77**	**63**	**84**	**36**	**129**	**948**

Female Voice

← Judge "Female" Judge "Male" →

"O/N"	1	2	3	4	5	6	7	8	9	10	11	12	13	Σ
1	99	35	66	48	48	38	85	38	38	43	35	12	31	**616**
2	1	1	11	7	16	14	23	14	12	9	8	0	0	**116**
3	3	0	1	4	7	7	15	3	7	1	0	0	2	**50**
4	0	2	2	1	3	3	24	3	5	0	1	0	1	**45**
5	1	1	1	1	0	2	31	6	1	0	2	1	0	**47**
6	0	0	0	0	0	4	19	1	1	0	0	0	0	**25**
7	1	0	0	0	1	1	42	0	1	0	0	0	0	**46**
Σ	**105**	**39**	**81**	**61**	**75**	**69**	**239**	**65**	**65**	**53**	**46**	**13**	**34**	**945**

New Items

← Judge "Female" Judge "Male" →

"O/N"	1	2	3	4	5	6	7	8	9	10	11	12	13	Σ
1	7	9	9	9	22	25	66	17	17	10	11	6	8	**216**
2	1	1	6	8	2	13	37	13	8	3	5	0	1	**103**
3	0	0	2	0	4	17	33	3	2	0	1	0	1	**65**
4	2	0	1	4	2	9	65	13	2	1	1	0	0	**101**
5	0	0	1	2	1	8	83	8	1	1	1	1	0	**107**
6	0	0	0	0	0	4	118	6	1	1	0	0	0	**130**
7	1	0	0	0	0	0	234	0	0	0	0	0	0	**235**
Σ	**11**	**10**	**19**	**23**	**31**	**76**	**636**	**60**	**31**	**24**	**19**	**7**	**10**	**957**

Note. "O/N" = "Old–New" confidence rating.

In both cases there is a two-parameter fit to the data, so it is possible to compare the chi-square values of the best-fit models.

A caveat must be added regarding this procedure for fitting ROC and zROC curves. In this procedure we have assumed that the false alarm rate (and its z score) is known exactly, so that one can perform a linear regression. The proper method is to perform a chi-square fit to both the hit rate and false alarm rate data. Furthermore, the cumulative probabilities are highly correlated, so

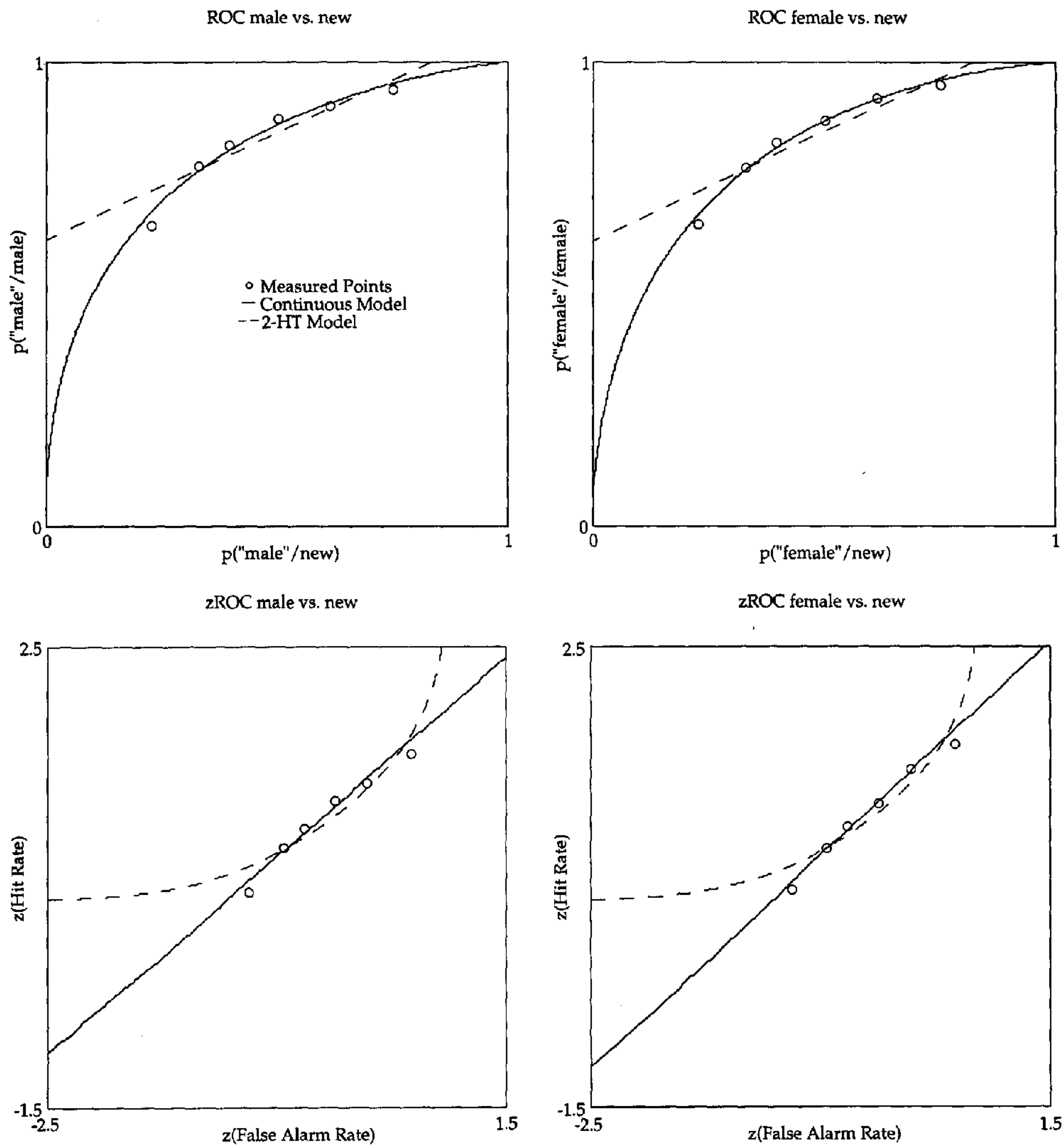

Figure 4. From Experiment 1, the old–new recognition receiver operating characteristic (ROC) and best-fit signal detection and two-high threshold models are shown for the male versus new data in the top-left figure and the female versus new data in the top-right figure. Corresponding z-transformed ROCs (zROCs) are shown in the bottom figures. 2-HT = two-high threshold.

the proper chi-square analysis should use the probability of each rating category rather than the cumulative probabilities (Levi & Klein, 1984; Macmillan & Creelman, 1991). However, we have used the linear regression method on the cumulative data (Equations 3 and 4) for comparison to other research in this field (Ratcliff et al., 1992; Yonelinas, 1999).

The signal detection model defines d' as distance from the origin to the *z*ROC multiplied by $\sqrt{2}$ as recommended by Macmillan and Creelman (1991). For old–new recognition, d' was 1.18 for both male items and female items. The ratio of the standard deviations for new to old items was .86 ± .15 for male items and .93 ± .12 for female items (95% confidence intervals reported). Both standard deviation ratios were near Ratcliff's proposed value of .8 (Ratcliff et al., 1992; 1994). In terms of confidence bias, the mean criterion value was −.47 and −.49 for male and female items, respectively. Thus, in terms of item recognition, standard deviation ratio, and criteria placement, studied items presented with a male voice were comparable with items presented with a female voice.

Figure 5 shows the ROCs and zROCs for source identification. The figures on the left were produced with data summed (collapsed) over all "old–new" ratings. Plotted with the data is the best-fit continuous model and best-fit, two-high threshold model. Collapsed source identification d' was .57, average criterion bias was $-.03$, and male–female standard deviation ratio was $.97 \pm .05$.

Because there is very little source information within the "old–new" ratings of 2–7, collapsing over all ratings when conducting source analysis has the effect of washing out the relevant source information. This is due to the addition of the large number of responses at or centered around "male–female" rating bin 7, where previously presented words are not remembered or word source is unknown. When collapsing the data, the ROC is "pulled" toward the chance line resulting in a flattened ROC. This artifactual flattening would be expected to result in a better fit for the threshold models and fair worse for the continuous model.

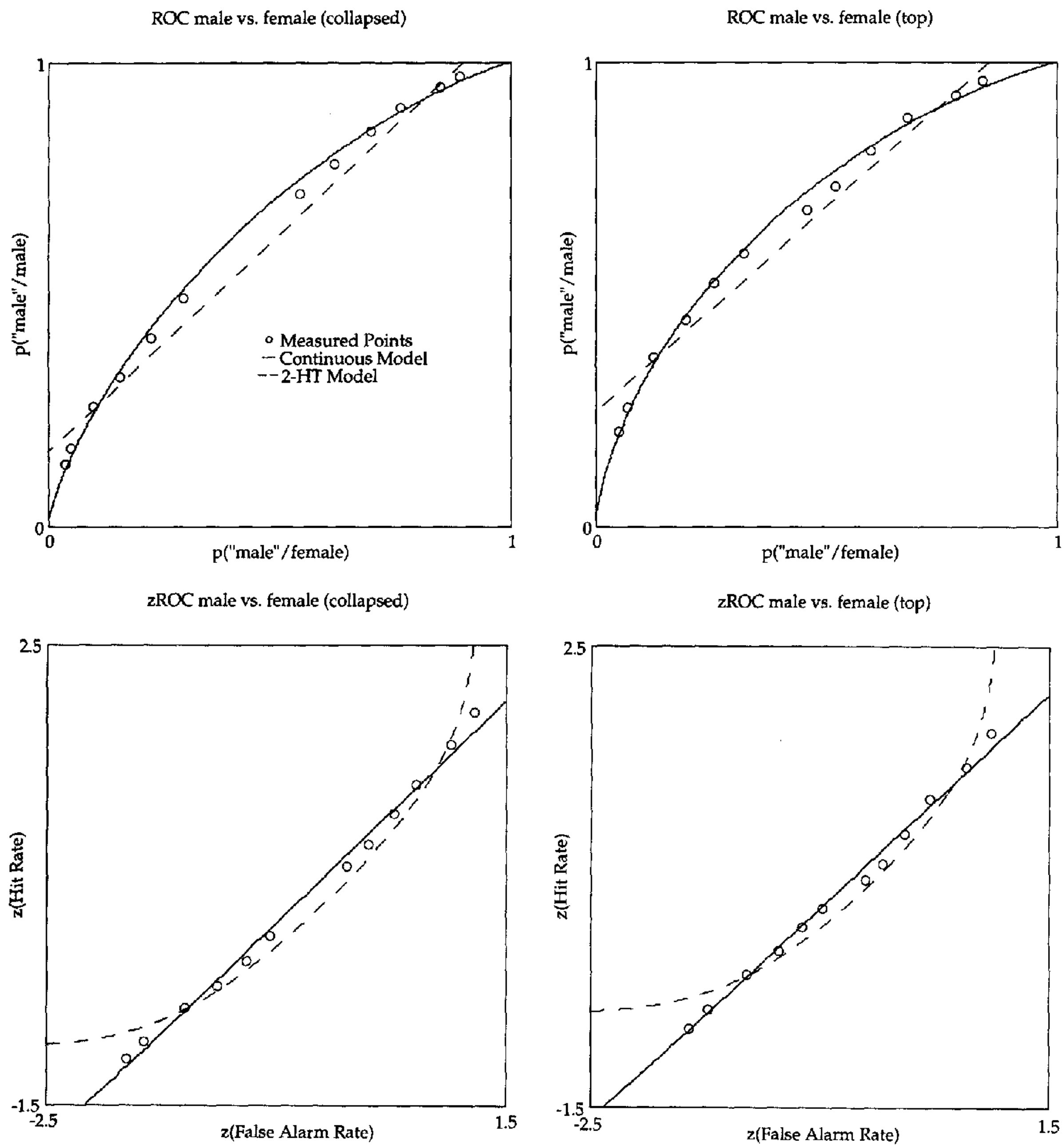

Figure 5. The collapsed and top source identification receiver operating characteristics (ROCs) and z-transformed ROCs (zROCs) from Experiment 1 with the best-fit models. 2-HT = two-high threshold.

To refine the source analysis, the male versus female condition was also conducted on data from the top "old–new" rating of 1 (Murnane & Bayen, 1996). The top source identification d' was .71, average criterion bias was −.04, and male–female standard deviation ratio was .93 ± .05. As the noise included in the analysis is reduced in the top rating condition, the increase in d' as compared to the collapsed condition is expected. The proximity of the standard deviation ratio to 1 is also expected because the memory strength distributions of the male source and female source should be approximately the same. In sum, the continuous model offered reasonable parameters that characterized item detection, source identification, criterion biases, and the variability of the underlying distributions.

Chi-square analysis. From the confidence ratings, we evaluated the ROCs using chi-square analysis for both item detection and source identification in terms of the degree to which they conformed generally to a continuous (curvilinear) or a two-high threshold (linear) model (see Table 3). For old–new judgments, the ROCs for item recognition were adequately fit by a continuous model and were not adequately fit by a two-high threshold model. Specifically, for the continuous model the fits for male items and female items were adequate, whereas the fits for the high-threshold models were significantly inadequate. To ensure these effects were not due to averaging over subjects, a subject-by-subject analysis was conducted using the maximum likelihood ROC analysis (Dorfman & Alf, 1969). Although it seldom occurred, subjects with perfect performance (i.e., hit rate = 1 at varying false alarm rates) were excluded from the analysis. For these subjects, both models fit the data perfectly, and thus its inclusion could not be used to distinguish between the models. The individual subject analyses produced similar results for both male (2 subjects excluded) and female (2 subjects excluded) items, though these comparisons failed to reach significance. In this case of item memory, we can only argue for the relative superiority of the continuous model. Anomalously low chi-square values could be due to the lack of independence of ROC points while anomalously high chi-square values could be due to chi-square fitting and summation in only the vertical dimension rather than in both the horizontal and vertical dimension.

Although the continuous model fit the ROC better, neither model fit the collapsed source ROC. Only the continuous model adequately fit the top source ROC. The individual subject source identification analysis confirmed the results given by the averaged data for both the collapsed data (no subjects excluded) and the top data (no subjects excluded). In the individual subject analyses, the continuous model fit was adequate for both source conditions. To ensure that the results were not an artifact of having more than one hit rate point for a false alarm rate of zero (which may have produced a hockey-stick shaped ROC favoring the continuous model), the individual subject analysis was also conducted excluding such subjects and the significance of all results remained the same (5 subjects excluded).

Linearity analyses. The divergent predictions of both the continuous model and two-high threshold models can also be indirectly tested using linearity analysis (Yonelinas, 1999). Specifically, the two-high threshold model predicts a linear ROC, and the continuous model predicts a linear zROC. In addition, the continuous model predicts an inverted U-shaped curvilinear ROC, whereas the two-high threshold model predicts a U-shaped curvilinear zROC. Linearity of the ROC or zROC can be tested by determining if the addition of a quadratic term provides a significantly better fit than a linear fit. Basically, linearity analysis consists of first fitting a line to either the ROC or zROC to determine whether there is a significant linear component in the function (rejecting the null hypothesis that a line provides an identical fit to the mean of the data) and then fitting a line + quadratic to the ROC or zROC to determine whether there is a significant quadratic component in the function (rejecting the null hypothesis that a line + quadratic provides an identical fit to a fit of a line alone). In testing the significance of the linear component, standard linear regression was conducted, treating the false alarm rates (x-axis direction) as fixed and regressing in the y-axis direction (i.e., the hit rate). The test for significance of the quadratic component was conducted twice: (1) by fixing the false alarm rates and regressing in the y-axis direction and (2) by fixing the hit rates and regressing in the x-axis direction. The data reported comes from the regression in either the x-axis or y-axis direction with the lowest sum of squared error (i.e., the best fit). Testing for linear significance was also conducted in both the x- and y-axis direction but did not affect any of the conclusions reported; therefore, we only report linear regression results in the y-axis direction. A significance level of $p < .05$ is used in all experiments. These procedures are identical to those used by Yonelinas (1999). Also reported is the quadratic component (c) from the regression in the y-axis direction, which can be used to determine the direction of curvature (i.e., U-shape or inverted U-shape). This is important in the analysis of zROC shapes (Glanzer et al., 1999b).

Linearity analysis (see Table 4) showed that recognition ROCs were curvilinear for both male items and female items. Similarly, both source conditions resulted in curvilinear ROCs. The curvilin-

Table 3
Chi-Square Analysis Results (Experiment 1)

ROC type	Group analysis: Continuous	Group analysis: High-threshold	Individual subject analysis: Continuous	Individual subject analysis: High-threshold
Male items	$\chi^2(4) = 9.31, p = \mathbf{.054}$	$\chi^2(4) = 41.86, p < .001$	$\chi^2(88) = 49.81, p = \mathbf{1.00}$	$\chi^2(88) = 110.66, p = \mathbf{.051}$
Female items	$\chi^2(4) = 5.40, p = \mathbf{.25}$	$\chi^2(4) = 41.87, p < .001$	$\chi^2(84) = 41.83, p = \mathbf{1.00}$	$\chi^2(84) = 94.21, p = \mathbf{.21}$
Collapsed source	$\chi^2(10) = 26.25, p = .0034$	$\chi^2(10) = 112.25, p < .001$	$\chi^2(240) = 213.05, p = \mathbf{.89}$	$\chi^2(240) = 385.93, p < .001$
Top source	$\chi^2(10) = 14.69, p = \mathbf{.14}$	$\chi^2(10) = 84.27, p < .001$	$\chi^2(240) = 184.06, p = \mathbf{1.00}$	$\chi^2(240) = 341.96, p < .001$

Note. Bold p values indicate an adequate fit. ROC = receiver operating characteristic.

Table 4
Linear Analysis Results (Experiment 1)

	R^2	F	MSE	c
ROC type				
Male items				
Linear	.8844	**F(1, 4) = 30.60**	.0016	
Quadratic	.9877	**F(1, 3) = 25.20**	.0002	−1.13
Female items				
Linear	.9026	**F(1, 4) = 37.09**	.0015	
Quadratic	.9933	**F(1, 3) = 40.28**	.0001	−1.09
Collapsed source				
Linear	.9875	**F(1, 10) = 789.71**	.0014	
Quadratic	.9994	**F(1, 9) = 187.31**	< .0001	−0.52
Top source				
Linear	.9728	**F(1, 10) = 357.64**	.0020	
Quadratic	.9975	**F(1, 9) = 88.06**	.0002	−0.65
zROC type				
Male items				
Linear	.9744	**F(1, 4) = 152.25**	.0058	
Quadratic	.9981	**F(1, 3) = 37.95**	.0008	−0.28
Female items				
Linear	.9845	**F(1, 4) = 253.42**	.0041	
Quadratic	.9983	**F(1, 3) = 23.80**	.0006	−0.24
Collapsed source				
Linear	.9936	**F(1, 10) = 1548.7**	.0074	
Quadratic	.9996	**F(1, 9) = 124.33**	.0006	0.09
Top source				
Linear	.9923	**F(1, 10) = 1296.3**	.0057	
Quadratic	.9977	**F(1, 9) = 20.55**	.0021	0.09

Note. Bold F values indicate a significant component. ROC = receiver operating characteristic; zROC = z-transformed ROC.

ear source ROCs are in accordance with the predictions of the continuous model and are not in accordance with the predictions of the two-high threshold model.

Linearity analysis of the zROCs showed that both item memory functions were curvilinear. Although the quadratic component significantly improved the fit of the data, the negative quadratic coefficients indicate the curves are inverted U-shapes. An inverted U-shape is not predicted by either the continuous model or two-high threshold model. Therefore, these curvilinear fits do not show support for either model. However, the negative coefficients do correspond with a predicted variability centered about zero shown by an analysis of recognition memory and used to support the continuous model predictions (Glanzer et al., 1999b). Both source zROCs were also shown to be curvilinear. The curvilinear zROCs are contradictory to the predictions of the continuous model and in line with the predictions of the threshold models. Thus, linearity analysis of the source ROCs provided evidence against the two-high threshold model and in support of the continuous model and linearity analysis of the source zROCs provided evidence against the continuous model and in support of the two-high threshold model.

Experiment 2

In Experiment 2, we used a different rating procedure to assess source memory. In the first experiment, source memory judgments were obtained by the integration of two separate confidence judgments (with a third judgment for old–new). That is, for each test item, we obtained one confidence rating that determined the degree to which participants judged the item as a Source A item and another rating that determined the degree to which participants judged the item as a Source B item. There are some advantages to the use of two separate judgments. First, there is no need to force participants to integrate both source judgments on a single dimension—for example, 1 = *high confidence that the item came from*

Table 5
Item Detection and Source Ratings (Experiment 2)

Male Voice

	← Judge "Female"				Judge "Male" →			
"O/N"	7	6	5	4	3	2	1	Σ
1	25	24	24	21	57	102	298	**551**
2	0	7	14	16	18	22	5	**82**
3	0	5	8	16	23	9	0	**61**
4	1	0	0	49	0	0	0	**50**
5	0	0	2	20	0	0	1	**23**
6	0	1	0	24	1	0	0	**26**
7	1	0	0	68	1	0	0	**863**
Σ	**27**	**37**	**48**	**214**	**100**	**133**	**304**	**863**

Female Voice

	← Judge "Female"				Judge "Male" →			
"O/N"	7	6	5	4	3	2	1	Σ
1	265	121	44	19	41	31	24	**545**
2	3	24	22	15	19	6	0	**89**
3	0	6	16	15	12	3	0	**52**
4	0	0	2	62	0	0	0	**64**
5	0	0	3	14	0	0	0	**17**
6	1	1	2	26	1	0	0	**31**
7	4	2	0	57	0	0	0	**63**
Σ	**273**	**154**	**89**	**208**	**73**	**40**	**24**	**861**

New

	← Judge "Female"				Judge "Male" →			
"O/N"	7	6	5	4	3	2	1	Σ
1	3	5	2	11	3	2	3	**29**
2	0	4	3	4	7	2	0	**20**
3	0	1	5	15	10	2	1	**34**
4	1	0	1	137	2	0	0	**141**
5	0	0	6	68	0	0	0	**74**
6	0	2	2	139	1	0	0	**144**
7	0	0	1	416	0	0	1	**418**
Σ	**4**	**12**	**20**	**790**	**23**	**6**	**5**	**888**

Note. "O/N" = "Old–New" confidence rating.

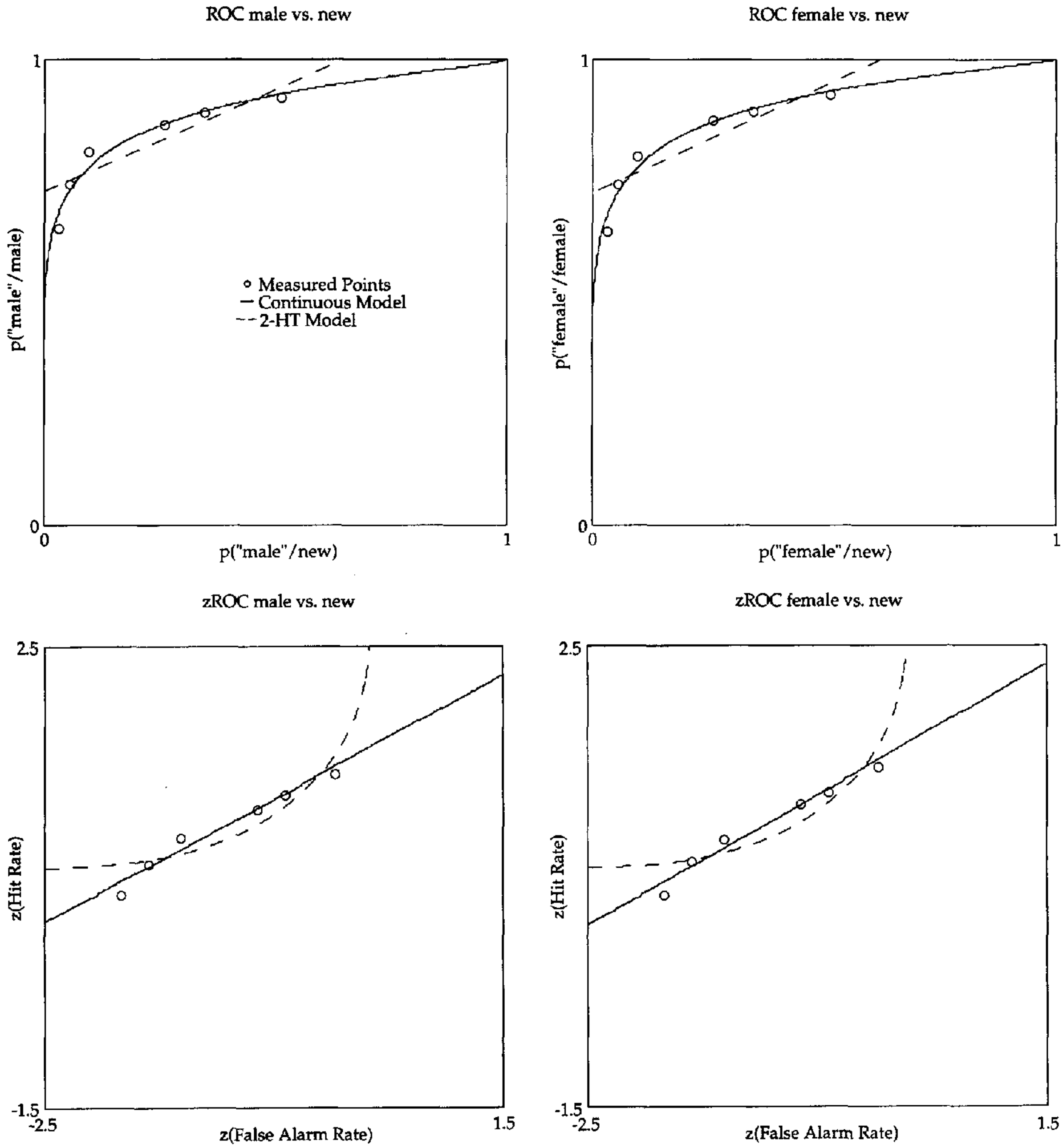

Figure 6. The old–new recognition receiver operating characteristics (ROCs) and *z*-transformed ROCs (*z*ROCs) from Experiment 2 with the best-fit models. 2-HT = two-high threshold.

Source A, and 7 = *high confidence that the item came from Source B*. Yet, this kind of two-judgment rating is not common and anomalies may have occurred when the data set from the two source judgments were integrated into one. Thus, in Experiment 2 we used a single confidence rating scale to assess source memory.

Method

Participants. The participants were 27 undergraduates from the University of California, Berkeley, who were each paid $8 for their participation.

Materials. The stimuli were similar to those used in Experiment 1. The target materials consisted of 96 nouns that were divided into three sets of 32 that were matched for length (5 letters) and frequency (M = 90) (Kučera & Francis, 1967). The three sets of words were rotated in the experimental design so that each set was spoken by the male voice and the female voice at study and also served as new items on the test. The study list contained 74 words, with the words in the first five and last five positions serving as buffer items. Of the remaining 64 target words, 32 were spoken by the male and 32 were spoken by the female. The words were randomly intermixed with the constraint that no more than three words from one voice appeared consecutively. After all 64 target words

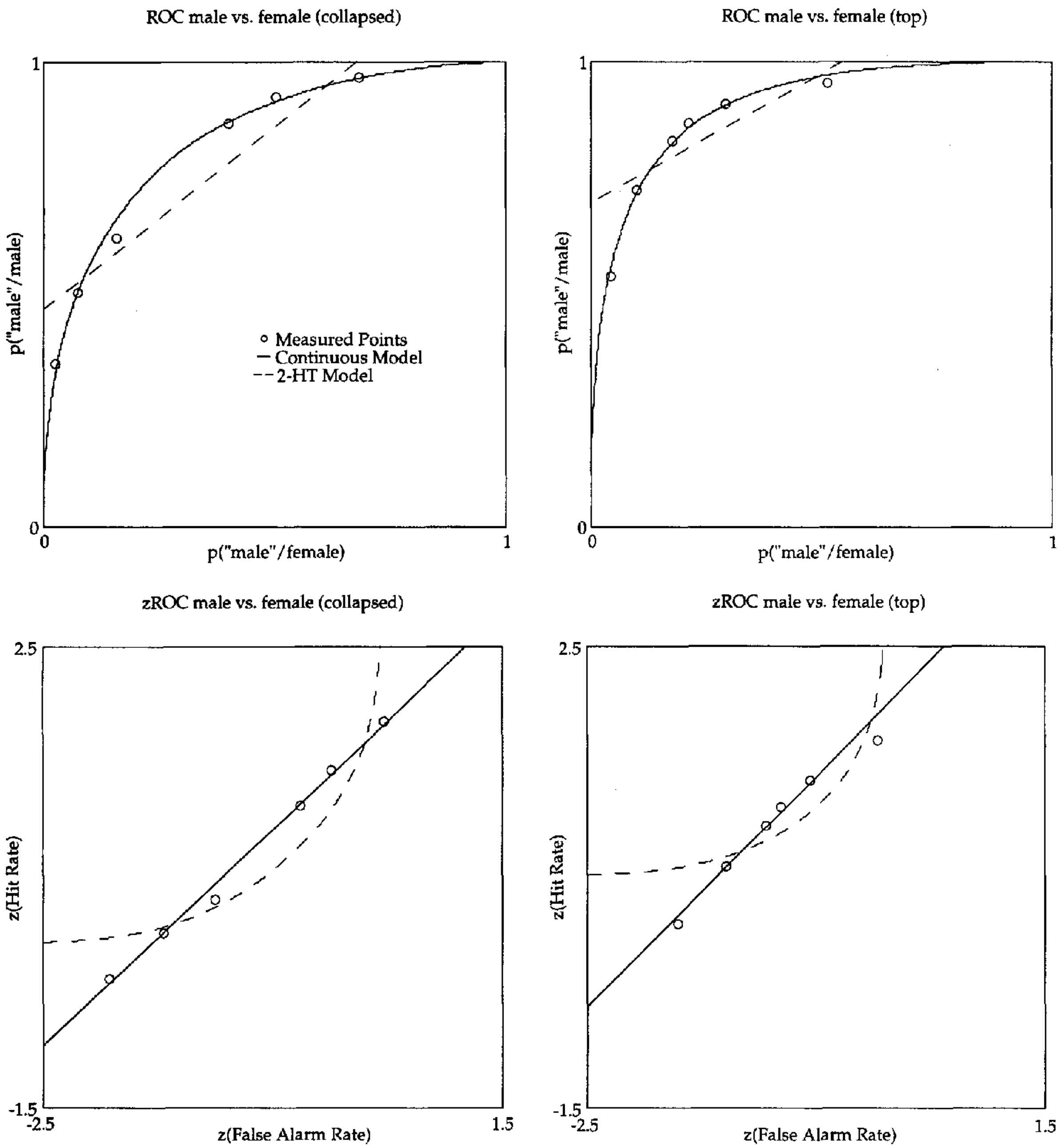

Figure 7. The collapsed and top source identification receiver operating characteristics (ROCs) and z-transformed ROCs (zROCs) from Experiment 2 with the best-fit models. 2-HT = two-high threshold.

were presented, they were repeated in a different random order. The visually presented test list consisted of 96 target words (64 old words and 32 new words) and an additional 10 practice words at the beginning of the list that were not scored.

Procedure. The study procedure was identical to that used in Experiment 1. After the study phase, participants were given the memory test and were informed that the test contained both new words and old words spoken by the male and the female. Participants were told that they would make two judgments for each word. First, they would rate their confidence (1–7) about whether the word was old or new (1 = *very confident "old;"* 7 = *very confident "new"*). Second, participants were instructed to rate their confidence (1–7) about the source of each word (1 = *very confident that the male spoke the word;* 7 = *very confident that the female spoke the word*). As in Experiment 1, participants were told to use any response on the 7-point scale that corresponded to their memory strength.

Results and Discussion

Continuous model parameter estimation. The rating distributions for all sources are given in Table 5 (responses other than 1–7 were eliminated). Figure 6 displays the old–new recognition ROCs and zROCs for both male and female item memory. Signal detection and high-threshold models were fit to the data. Male items had a d' of 1.80 and female items had a d' of 1.82. Criteria were .01 for both male and female items. The standard deviation ratio was .54 ± .12 for male items and .57 ± .11 for female items. These values deviate from Ratcliff et al.'s (1992) constant of .8 and show that old–new variance is indeed variable as shown by Glanzer et al. (1999b). As in Experiment 1, recognition memory strength, mean critia location, and standard deviation ratio were similar for male and female items.

The collapsed and top source identification ROCs and zROCs are illustrated in Figure 7. For the collapsed data, d' was 1.41, criterion was at −.02, and the standard deviation ratio was .94 ± .08. For the top data, d' was 1.86, criterion was at −.05 and standard deviation ratio was 1.01 ± .13. As in Experiment 1, collapsing over "old–new" ratings resulted in a lower d' because of the addition of noise. In addition, the proximity of the variability ratio to unity indicates that the distribution of source memory strength was similar.

Chi-square analysis. For item detection, the continuous model for male items and female items did not adequately fit the ROCs, but they still fared better numerically than the two-high threshold model (see Table 6). Only the continuous model provided an adequate fit in the individual subject analysis for both male (one subject removed) and female items (no subjects removed).

As in Experiment 1, neither model fit the collapsed source data, whereas only the continuous model fit the top source data. The individual subject analysis resulted in an adequate fit by the continuous model for both source conditions. To ensure that the results were not an artifact of having more than one hit rate point for a false alarm rate of zero, the individual subject analysis was also conducted excluding such subjects and the significance of all results remained the same (3 subjects excluded). The findings of using a single rating for source memory were generally consistent with the findings of Experiment 1 and thus indicate that the results obtained in that experiment were not an artifact of the three-judgment source memory procedure.

Linearity analysis. Recognition ROCs were curvilinear for both male items and female items (see Table 7). Both source conditions resulted in curvilinear ROCs as well.

In contrast with Experiment 1, linearity analysis of the zROCs showed that both item memory functions were linear. The collapsed source memory zROC was also shown to be linear, whereas the top source memory zROC was shown to be curvilinear. However, the negative quadratic component indicates the top zROC is an inverted-U shape, which does not correspond well with the predictions of either model. Like the negative values of c reported by Glanzer et al. (1999b) in item memory, this negative quadratic component in source identification could similarly be interpreted as a variation about a mean c of zero predicted by the continuous model. Overall, the linearity analysis of the source ROCs provided evidence in favor of the continuous model and provided evidence against the two-high threshold model. Unlike Experiment 1, linearity analysis of the collapsed source zROC also provided evidence in favor of the continuous model and provided evidence against the two-high threshold model while linearity analysis of the top source zROC provided evidence which did not support the predictions of either model.

Table 7
Linear Analysis Results (Experiment 2)

	R^2	F	MSE	c
ROC type				
Male items				
Linear	.7862	$F(1, 4)$ = **14.71**	.0030	
Quadratic	.9827	$F(1, 3)$ = **34.14**	.0010	−1.39
Female items				
Linear	.7983	$F(1, 4)$ = **15.83**	.0031	
Quadratic	.9864	$F(1, 3)$ = **41.46**	.0008	−1.43
Collapsed source				
Linear	.9152	$F(1, 4)$ = **43.19**	.0067	
Quadratic	.9911	$F(1, 3)$ = **25.51**	.0009	−1.56
Top source				
Linear	.7079	$F(1, 4)$ = **9.69**	.0011	
Quadratic	.9705	$F(1, 3)$ = **26.69**	.0011	−3.12
zROC type				
Male items				
Linear	.9627	$F(1, 4)$ = **103.25**	.0071	
Quadratic	.9837	$F(1, 3)$ = 3.86	.0041	−0.16
Female items				
Linear	.9724	$F(1, 4)$ = **140.82**	.0059	
Quadratic	.9876	$F(1, 3)$ = 3.69	.0035	−0.15
Collapsed source				
Linear	.9943	$F(1, 4)$ = **703.53**	.0055	
Quadratic	.9971	$F(1, 3)$ = 2.77	.0038	0.08
Top source				
Linear	.9765	$F(1, 4)$ = **166.35**	.0092	
Quadratic	.9984	$F(1, 3)$ = **41.40**	.0008	−0.23

Note. Bold F values indicate a significant component. ROC = receiver operating characteristic; zROC = z-transformed ROC.

Table 6
Chi-Square Analysis Results (Experiment 2)

	Group analysis		Individual subject analysis	
ROC type	Continuous	High-threshold	Continuous	High-threshold
Male items	$\chi^2(4)$ = 12.98, p = .011	$\chi^2(4)$ = 60.49, p < .001	$\chi^2(104)$ = 71.21, p = **.99**	$\chi^2(104)$ = 187.47, p < .001
Female items	$\chi^2(4)$ = 10.85, p = .028	$\chi^2(4)$ = 62.69, p < .001	$\chi^2(108)$ = 91.23, p = **.88**	$\chi^2(108)$ = 182.95, p < .001
Collapsed source	$\chi^2(4)$ = 10.04, p = .040	$\chi^2(4)$ = 218.80, p < .001	$\chi^2(104)$ = 67.73, p = **1.00**	$\chi^2(104)$ = 280.35, p < .001
Top source	$\chi^2(4)$ = 5.47, p = **.24**	$\chi^2(4)$ = 120.02, p < .001	$\chi^2(100)$ = 62.03, p = **1.00**	$\chi^2(100)$ = 241.77, p < .001

Note. Bold p values indicate an adequate fit. ROC = receiver operating characteristic.

Experiment 3

Many participants in the previous experiments performed very well (perfectly in a few cases). It is possible that the previous results may have been influenced by abnormally high performance levels. To reduce the level of performance, this experiment utilizes a longer study list. This also has the fortunate side-effect of increasing the statistical power in the analysis.

Method

Participants. The participants were 24 undergraduates from Harvard University, who were each paid $8 for their participation.

Materials. Materials were similar to those used in Experiment 2, but 80 words were used for each item type rather than 32 words.

Procedure. The procedure was identical to Experiment 2 except for a reversal in the rating scales (for old–new ratings, 1 = *very confident "new"* and 7 = *very confident "old;"* for source ratings, 1 = *very confident "female"* and 7 = *very confident "male"*).

Results and Discussion

Continuous model parameter estimation. Table 8 contains the rating distributions for all item types (responses other than 1–7 were eliminated). Figure 8 displays the old–new recognition ROCs and *z*ROCs for both male and female item detection. Male items had a d' of 1.59 and female items had a d' of 1.57. Criteria were placed at −.12 and −.11 for both male and female items, respectively. The standard deviation ratio was .76 ± .04 for male items and .75 ± .04 for female items. These values are close to Ratcliff et al.'s (1992) constant of .8. As in the previous experiments, parameter values are similar for both male and female items.

The collapsed and top source identification ROCs and *z*ROCs are illustrated in Figure 9. For the collapsed data, d' was 1.08, criterion was at .02, and male–female standard deviation ratio was .98 ± .21. The top data resulted in a d' of 1.66, a criterion placement of .01, and a standard deviation ratio of .98 ± .11. As in the previous experiments, the d' from the collapsed data is lower than the d' from the top data and the standard deviation ratios are near unity. By comparing the right and left figures in the top of Figure 9, one can see that collapsing the data effectively pulls the ROC toward the chance line, thereby lowering d'.

Chi-square analysis. For item detection, only the continuous model adequately fit the data (see Table 9). The individual subject analysis produced the same results (no subjects excluded).

Although the continuous model fit the data better for both source memory conditions, neither model adequately fit the data. The results were confirmed by the individual subject analysis in the collapsed condition (no subjects excluded); however, individual subject analysis did show an adequate fit for the continuous model in the top source condition (no subjects excluded). To ensure that the results were not an artifact of having more than one hit rate point for a false alarm rate of zero, the individual subject analysis was also conducted excluding such subjects, and the significance of all results remained the same (one subject excluded). The lower d's in all conditions and lack of exclusion of subjects because of perfect performance in this experiment, as compared to Experiment 2, indicate that performance was not near ceiling.

Table 8
Item Detection and Source Ratings (Experiment 3)

Male Voice

	← Judge "Female"				Judge "Male" →			
"O/N"	1	2	3	4	5	6	7	Σ
7	22	73	86	49	103	202	385	920
6	1	49	57	41	102	77	7	334
5	0	16	86	70	82	5	0	259
4	0	0	3	98	1	2	0	104
3	0	1	8	117	6	0	0	132
2	0	2	7	106	4	0	0	119
1	0	0	0	49	2	0	0	51
Σ	23	141	247	530	300	286	392	1919

Female Voice

	← Judge "Female"				Judge "Male" →			
"O/N"	1	2	3	4	5	6	7	Σ
7	399	181	114	55	74	63	24	910
6	8	89	77	40	78	36	2	330
5	0	22	94	77	68	11	0	272
4	1	0	0	103	0	2	0	106
3	0	0	7	109	9	0	0	125
2	0	1	9	106	3	0	0	119
1	0	0	0	58	0	0	0	58
Σ	408	293	301	548	232	112	26	1920

New

	← Judge "Female"				Judge "Male" →			
"O/N"	1	2	3	4	5	6	7	Σ
7	5	6	17	6	12	8	1	55
6	0	20	21	26	30	12	0	109
5	0	10	75	87	63	4	0	239
4	0	0	10	157	2	0	0	169
3	0	2	15	321	14	1	0	353
2	0	1	12	494	8	0	0	515
1	0	1	1	475	2	0	0	479
Σ	5	40	151	1566	131	25	1	1919

Note. "O/N" = "Old–New" confidence rating.

Linearity analysis. Linearity analysis showed that recognition ROCs were curvilinear for both male items and female items (see Table 10). Both source conditions also resulted in curvilinear ROCs. By comparing the quadratic term for the collapsed data with that of the top data in this experiment (−.84 vs. −2.27), it is apparent that collapsing the data has the effect of flattening out the ROC.

Linearity analysis of the *z*ROCs showed that both item memory functions were linear, whereas both source memory functions were

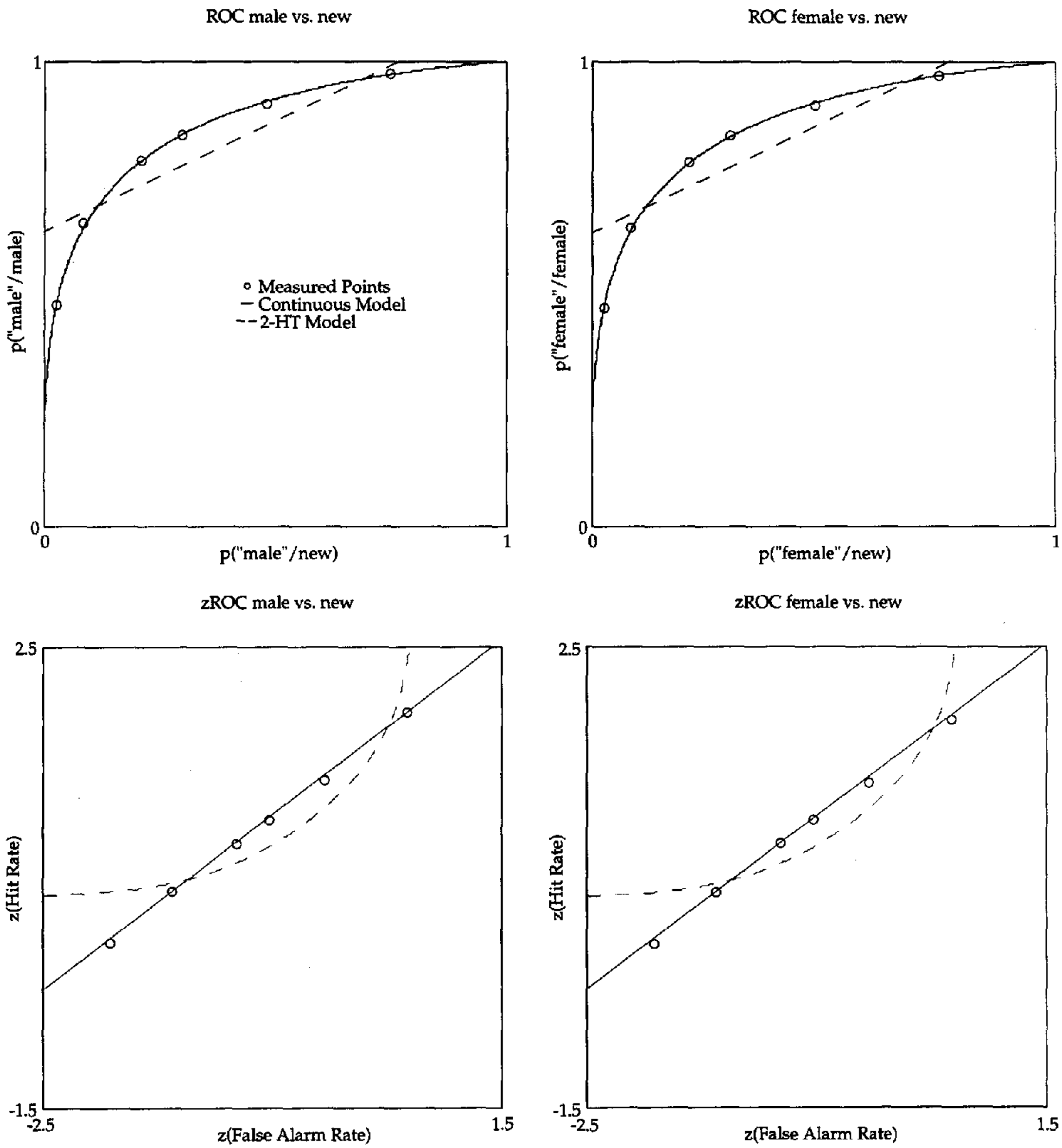

Figure 8. The old–new recognition receiver operating characteristics (ROCs) and z-transformed ROCs (zROCs) from Experiment 3 with the best-fit models. 2-HT = two-high threshold.

curvilinear. Therefore, the linearity analysis of the source ROCs shows support for the continuous model and provides evidence against the two-high threshold model, whereas the linearity analysis of the source zROCs shows support for the two-high threshold model and provides evidence against the continuous model.

General Discussion

In all three experiments, linearity analyses of the source ROCs provided evidence in favor of the continuous model and against the two-high threshold model. Linearity analyses of the source zROCs provided mixed results: 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. The results of the linearity analysis do not show that either model is correct; however, the threshold model was rejected in all cases by the source ROC analyses, and the continuous model was not rejected in all

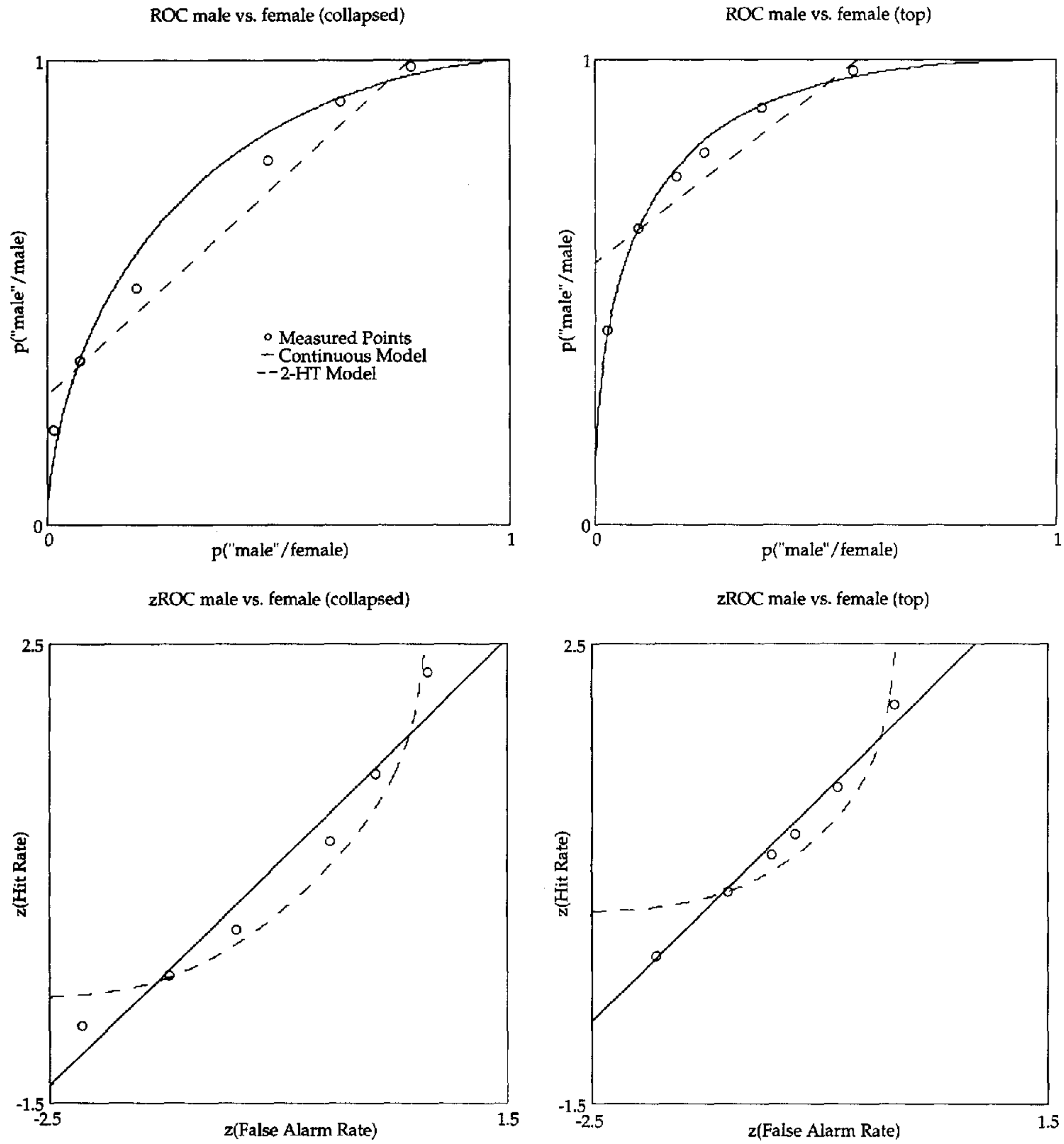

Figure 9. The collapsed and top source identification receiver operating characteristics (ROCs) and z-transformed ROCs (zROCs) from Experiment 3 with the best-fit models. 2-HT = two-high threshold.

cases by the source zROC analyses. It is also important to remember that linearity analysis is an *indirect* test of the predictions of the models. The fit of a quadratic term is not identical to the curvature of the ROC predicted by the continuous model or the curvature of the zROC predicted by the two-high threshold model. Although useful as a convergent method of analysis, the results of the linearity analysis do not comprise an actual test of the models themselves. In contrast, the chi-square analysis is conducted by fitting each model to the data and then determining the error of this fit. Thus, the chi-square analysis is a *direct* measure of model adequacy.

In all cases of the chi-square analysis, the continuous model fit the data numerically better than a two-high threshold model. In fact, the threshold model did not adequately fit the data in any case. This suggests that a two-high threshold model is not an appropriate model of source memory in general. The result that our model did not always adequately fit the data does not surprise us, as it seems unlikely that the complex cognitive process of source

Table 9
Chi-Square Analysis Results (Experiment 3)

	Group analysis		Individual subject analysis	
ROC type	Continuous	High-threshold	Continuous	High-threshold
Male items	$\chi^2(4) = 1.55$, $p =$ **.82**	$\chi^2(4) = 397.90$, $p < .001$	$\chi^2(96) = 85.5$, $p =$ **.77**	$\chi^2(96) = 569.25$, $p < .001$
Female items	$\chi^2(4) = 1.99$, $p =$ **.74**	$\chi^2(4) = 410.80$, $p < .001$	$\chi^2(96) = 87.1$, $p =$ **.73**	$\chi^2(96) = 595.05$, $p < .001$
Collapsed source	$\chi^2(4) = 206.28$, $p < .001$	$\chi^2(4) = 262.23$, $p < .001$	$\chi^2(96) = 142.65$, $p = .0014$	$\chi^2(96) = 354.65$, $p < .001$
Top source	$\chi^2(4) = 12.78$, $p = .012$	$\chi^2(4) = 176.77$, $p < .001$	$\chi^2(96) = 74.36$, $p =$ **.95**	$\chi^2(96) = 269.89$, $p < .001$

Note. Bold p values indicate an adequate fit. ROC = receiver operating characteristic.

memory can be fit by any two-parameter model. In the world of modeling, the goal is to use the best-fitting model until a better model is created. It is surprising that in all cases where source memory ROCs were not influenced by the addition of noise (by collapsing the data) or anomalous averaging effects (by averaging over subjects), the continuous model adequately fit the data.

Both Donaldson, MacKenzie, and Underhill (1996) and Yonelinas (1999) applied signal detection methodology to item detection and to source identification. In terms of item memory, Yonelinas (1994) proposed a dual-process model of recognition memory in which memory performance is based on two independent judgments—a continuous judgment of item familiarity and an added threshold recollection decision. Recollection is presumed to involve a discrete process that increases recognition performance for items associated with strong source or episodic knowledge. Yonelinas (1994) showed that the dual-process model could account for various properties of the ROC and zROC curves assessed in his experiments on recognition memory. In particular, Yonelinas's findings added evidence against a simple signal detection model in which it is assumed that variance of new and old distributions are the same (i.e., $\sigma_n/\sigma_o = 1$). However, his data also appeared to be consistent with a signal detection model in which the variability ratio is allowed to vary freely (i.e., $\sigma_n \neq \sigma_o$). A recent analysis of a large body of recognition memory data provides evidence against the dual-process model and in support of the unequal variance signal detection model (Glanzer et al., 1999a, 1999b).

Table 10
Linear Analysis Results (Experiment 3)

	R^2	F	MSE	c
ROC type				
Male items				
Linear	.8019	$F(1, 4) =$ **16.19**	.0070	
Quadratic	.9839	$F(1, 3) =$ **33.92**	.0019	−1.24
Female items				
Linear	.7971	$F(1, 4) =$ **15.71**	.0085	
Quadratic	.9811	$F(1, 3) =$ **29.27**	.0023	−1.27
Collapsed source				
Linear	.9712	$F(1, 4) =$ **134.88**	.0036	
Quadratic	.9984	$F(1, 3) =$ **51.43**	.0003	−0.84
Top source				
Linear	.8442	$F(1, 4) =$ **21.67**	.0077	
Quadratic	.9918	$F(1, 3) =$ **54.35**	.0005	−2.27
zROC type				
Male items				
Linear	.9993	$F(1, 4) =$ **5643.3**	.0004	
Quadratic	.9993	$F(1, 3) = .0017$	.0006	< 0.01
Female items				
Linear	.9991	$F(1, 4) =$ **4290.7**	.0005	
Quadratic	.9995	$F(1, 3) = 2.58$	.0004	−0.02
Collapsed source				
Linear	.9519	$F(1, 4) =$ **79.21**	.0802	
Quadratic	.9992	$F(1, 3) =$ **173.22**	.0018	0.25
Top source				
Linear	.9843	$F(1, 4) =$ **250.94**	.0111	
Quadratic	.9990	$F(1, 3) =$ **44.28**	.0009	0.17

Note. Bold F values indicate a significant component. ROC = receiver operating characteristic; zROC = z-transformed ROC.

In terms of source memory, both Donaldson and MacKenzie (1996) and Yonelinas (1996) reported that ROC curves for source memory were generally linear. According to the dual-process model, if the familiarity of sources is approximately equal, then source identification will rely on recollection and ROCs will be linear. Yonelinas (1999) used a two-high threshold linear model (i.e. the restricted dual-process model) in three experiments where familiarity was assumed to play a smaller role than recollection in source identification. In the same study, a familiarity component was used in a fourth experiment, where source familiarity differences were created by presenting one list of words 5 days after the other, thereby increasing the familiarity of the more recently presented list. In all of our experiments, familiarity can be assumed to be approximately equal; however, the source memory ROCs were not linear contradicting the predictions of the dual-process model.

There may be several factors that determine the degree to which ROC curves suggest threshold functions. The first may be the effect of collapsing over "old–new" response ratings. Because the procedure used to assess source memory by Yonelinas (1999) utilizes a one-dimensional source rating scale, the responses are collapsed over "old–new." In addition, linear functions may be obtained if ratings are based on very limited information or features of an episode. For example, source memory performance may be based on requiring individuals to attend to a specific detail or single feature. In some cases, it may be that ratings about the presence or absence of that specific feature in memory is based on a discrete threshold judgment (i.e., either the individual remembers it or not). Another possibility is that source knowledge comes from a small subset of items that are remembered extremely well. This effect could disproportionately increase performance at the endpoints of the ROC curve and suggest a more linear function.

As indicated by the present analysis, a fairly standard signal detection approach to source memory performance showed that source memory is better characterized as a continuous rather than a discrete function. In another study (Dodson et al., 1998) source recollection was shown to involve partial source information in addition to specific source knowledge. In that study, participants were asked to recognize items presented by specific voices (e.g., two male voices, two female voices). Participants were able to discriminate among the specific sources (i.e., voices); however, when a source error was made, participants often chose a voice of the same gender as the correct voice. This finding suggests that partial information (i.e., correct gender, wrong person) contributed to source memory performance. In everyday experiences, it is likely that such partial knowledge mediates source recollection. Along with the present data, the finding that partial source information contributes to source memory judgments is also consistent with a continuous model of source memory.

References

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: I. Learning, memory and thinking* (pp. 243–293). San Francisco: W. H. Freeman.

Banks, W. P. (1970). Signal detection theory and human memory. *Psychological Bulletin, 74,* 81–99.

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

Batchelder, W. H., & Riefer, D. M. (1990). Multinomial processing models of source monitoring. *Psychological Review, 97,* 548–564.

Batchelder, W. H., Riefer, D. M., & Hu, X. (1994). Measuring memory factors in source monitoring: Reply to Kinchla. *Psychological Review, 101,* 172–176.

Bayen, U. J., Murnane, K., & Erdfelder, E. (1996). Source identification, item detection, and multinomial models of source monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 197–215.

Bernbach, H. A. (1967). Decision processes in memory. *Psychological Review, 74,* 462–480.

Buchner, A., Erdfelder, E., Steffens, M. C., & Martensen, H. (1997). The nature of memory processes underlying recognition judgments in the process dissociation procedure. *Memory & Cognition, 25,* 508–517.

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

Dodson, C. S., Prinzmetal, W., & Shimamura, A. P. (1998). Using Excel to estimate parameters from observed data: An example from source memory data. *Behavior Research Methods, Instruments, and Computers, 30,* 517–526.

Dodson, C. S., & Shimamura, A. P. (2000). Differential effects of cue dependency on item and source memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 1023–1044.

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

Donaldson, W., & MacKenzie, T. M. (1996, November). *Source memory operating characteristics.* Poster presented at the 37th Annual Meeting of the Psychonomic Society.

Donaldson, W., MacKenzie, T. M., & Underhill, C. F. (1996). A comparison of recollective memory and source monitoring. *Psychonomic Bulletin & Review, 3,* 486–490.

Dorfman, D. D., & Alf, E. (1969). Maximum-likelihood estimation of parameters of signal-detection theory and determination of confidence intervals: Rating-method data. *Journal of Mathematical Psychology, 6,* 487–496.

Drake, A. I., & Hannay, H. J. (1992). Continuous recognition memory tests: Are the assumptions of the theory of signal detection met? *Journal of Clinical & Experimental Neuropsychology, 14,* 539–544.

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

Erdfelder, E., & Buchner, A. (1998). Process-dissociation measurement models: Threshold theory or detection theory? *Journal of Experimental Psychology: General, 127,* 83–96.

Gardiner, J. M. (1988). Functional aspects of recollective experience. *Memory & Cognition, 16,* 309–313.

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

Glanzer, M., Hilford, A., Kim, K., & Adams, J. K. (1999a). Further tests of dual-process theory: A reply to Yonelinas (1999). *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 522–523.

Glanzer, M., Kim, K., Hilford, A., & Adams, J. K. (1999b). 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.

Hintzman, D. L. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review, 93,* 411–428.

Hirst, W. (1982). The amnesic syndrome: Descriptions and explanations. *Psychological Bulletin, 91,* 435–460.

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

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

Johnson, M. K., Kounios, J., & Reeder, J. A. (1994). Time-course studies of reality monitoring and recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 1409–1419.

Klein, S. A. (1985). Double-judgment psychophysics: problems and solutions. *Journal of the Optical Society of America, 2,* 1560–1585.

Kučera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English.* Providence, RI: Brown University Press.

Levi, D. M., & Klein, S. A. (1984). Detection and discrimination of the direction of motion in central and peripheral vision of normal and amblyopic observers. *Vision Research, 24,* 789–800.

Macmillan, N. A., & Creelman, C. D. (1991). *Detection theory: A user's guide.* New York: Cambridge University Press.

Mayes, A. R., Meudell, P. R., & Pickering, A. (1985). Is organic amnesia caused by a selective deficit in remembering contextual information? *Cortex, 21,* 167–202.

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

Murnane, K., & Bayen, U. J. (1996). An evaluation of empirical measures of source identification. *Memory & Cognition, 24,* 417–428.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. (1988). *Numerical recipes in C.* New York: Cambridge University Press.

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

Riefer, D. M., Hu, X., & Batchelder, W. H. (1994). Response strategies in source monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 680–693.

Schacter, D. L., Harbluk, J. L., & McLachlan, D. R. (1984). Retrieval without recollection: An experimental analysis of source amnesia. *Journal of Verbal Learning and Verbal Behavior, 23,* 593–611.

Shimamura, A. P., & Squire, L. R. (1987). A neuropsychological study of fact memory and source amnesia. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13,* 464–473.

Slotnick, S. D. (1996). *Recognition memory for source: A multidimensional signal detection analysis.* Unpublished master's thesis, University of California, Berkeley.

Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General, 117,* 34–50.

Tulving, E. (1972). Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), *Organization of memory* (pp. 381–403). New York: Academic Press.

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. (1996, October). *The contribution of recollection and familiarity to recognition and source memory: A formal model and an ROC analysis.* Paper presented at the 38th Annual Meeting of the Psychonomic Society, Chicago.

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

Zaragoza, M. S., & Lane, S. M. (1994). Source misattribution and the suggestibility of eyewitness memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 934–945.

Received August 19, 1997
Revision received February 15, 2000
Accepted February 25, 2000 ■