Journal of Experimental Psychology: General
2002, Vol. 131, No. 4, 494–510

0096-3445/02/$5.00 DOI: 10.1037//0096-3445.131.4.494

Regularities of Source Recognition: ROC Analysis

Andy Hilford and Murray Glanzer
New York University

Kisok Kim
Iona College

Lawrence T. DeCarlo
Columbia University

Source memory has become the focus of a growing number of investigations in a variety of fields. An appropriate model for source memory is, therefore, of increasing importance. A simple 2-dimensional signal-detection model of source recognition is presented. The receiver operating characteristics (ROCs) obtained from 3 experiments are then used to test the model. The data demonstrate 3 regularities: convex ROCs, *z*-ROCs with linear slopes of 1.00, and slightly concave *z*-ROCs. Two of these regularities support the model. The 3rd requires a revision of the model. This revised model is fitted to the data. The implications of these regularities for other theories are also discussed.

In this article we present data to evaluate formal[1] models of source recognition. With the growing use of source memory tests in several areas of psychology (normal adult memory, self-monitoring, psychopathology, psychoneurology, psychology of aging, developmental psychology, and analysis of bilingual performance), the need for evaluation is increased. Moreover, a decision about the proper form of theory for this area is not an academic exercise without practical implications. Kinchla (1994) demonstrated that analyzing behavior (e.g., whether two individuals differ in a basic capacity) depends on the theory used to evaluate the individuals' performance. He showed, with data from thought-disordered and non-thought-disordered patients (Harvey, 1985), that a threshold model leads to a diagnosis of a difference in capacity between the two groups, whereas a signal-detection model leads to a diagnosis of a simple bias difference.

The data of three experiments are presented and analyzed in this article. The analysis supports the existence of three regularities. These regularities are used to evaluate several models of source recognition: a signal-detection model, a revision of that model, a dual-process model, and a threshold model derived from current multinomial models. These are discussed in the final section of this article.

Andy Hilford and Murray Glanzer, Department of Psychology, New York University; Kisok Kim, Department of Psychology, Iona College; Lawrence T. DeCarlo, Teachers College, Columbia University.

The research reported in this article was carried out under National Institute of Mental Health Grant 1R01 MH60114. We thank the following people: Ingrid Kessler, who tested the participants in Experiments 1 and 3, for her excellent work in the laboratory; John Swets, for furnishing relevant background information; William Banks and Andy Yonelinas, for sending preprints of their articles; and, particularly, Pat Shrout and Larry Maloney for their advice on statistics and analysis.

Correspondence concerning this article should be addressed to Andy Hilford or Murray Glanzer, Department of Psychology, New York University, 6 Washington Place, New York, New York, 10003. E-mail: Andy Hilford at ah13@nyu.edu or Murray Glanzer at mg@psych.nyu.edu

The following topics are covered in the present article: (a) signal-detection theory and the receiver operating characteristic (ROC), (b) background on source recognition memory, (c) a signal-detection model of source recognition memory, (d) testing the model using three experiments, (e) revision of the signal-detection model and test of the revision, and (f) alternative models. The literature on signal-detection theory distinguishes two types of performance: detection and identification (Green & Swets, 1974; Macmillan & Creelman, 1991; Tanner, 1956). In cases of sensory discrimination, detection is determining whether or not a signal has been presented. Identification is determining which of two or more signals (e.g., a high-frequency or a low-frequency tone) has been presented. A full discussion of signal-detection theory applied to identification is found in Macmillan and Creelman (1991). The distinction between detection and identification parallels the distinction in memory literature between item recognition and source recognition. In item recognition, the participant judges whether a test item has occurred in a study list. In source recognition, the participant judges whether a test item came from one source or another (e.g., was it read by A or by B?). The concepts of signal-detection theory and the analytic procedures derived from it have been applied extensively to item recognition.

Signal-Detection Theory and the ROC

A key analytic procedure in signal-detection theory is the ROC. In the case of item recognition, this is a function relating hits (correct responses to old items as old) to false alarms (incorrect responses to new items as old) at various levels of bias or at

[1] Formal models are those in mathematical form. Models in mathematical form permit prediction of the shape of the receiver operating characteristic (ROC) and *z*-ROC. We use the term *theory* to refer to general approaches (e.g., signal-detection theory, threshold theory). We use the term *model* to refer to specific realizations of those approaches (e.g., normal signal-detection model, multinomial model).

various levels of confidence. The shape of the ROC tells whether a model based on signal-detection theory is appropriate for the performance. For example, signal-detection theory predicts convex ROCs when the usual assumption of underlying normal distributions is made.

There are two methods used to generate ROCs (Swets, Tanner, & Birdsall, 1961). One is a binary or fixed bias level method in which participants are induced to adopt different bias levels. This could be done, for example, in item recognition by giving different proportions of old and new test items or by varying rewards for hits and penalties for false alarms in different groups. The other method is based on confidence ratings (e.g., from 1 to 6) in which participants indicate their confidence rating for responses to each of the test items. The different confidence ratings are considered the equivalent of bias levels. The two methods, binary and confidence rating, give concordant information. Most important for our purposes, they produce ROCs that have the same general shape. In the case of item recognition, both methods produce, as expected on the basis of signal-detection theory, convex ROCs. Ratcliff, Sheu, and Gronlund (1992) used both binary and rating procedures for item recognition—binary in their Experiments 1 and 2, ratings in their Experiment 3. Both procedures produce convex ROCs. (This can be seen when the *z*-ROCs they present in their Figures 2, 3, and 5 are returned to their original untransformed shapes.)

In the work described here, we use confidence rating ROCs. A major part of the recognition memory literature is based on such ROCs. For example, there are 30 experiments summarized in Glanzer, Kim, Hilford, and Adams (1999) that use confidence rating ROCs.

Equation 1 describes a confidence rating ROC for item recognition. The ROC is a function relating the cumulative ratings of old items to the cumulative ratings of new items.

$$P(R_j|\text{old}) = f[P(R_j|\text{new})], \tag{1}$$

where R_j is a rating from confidence level j up to and including the highest, that is, strictest, rating.[2]

An ROC for source recognition memory has a homologous form. It is a function relating the confidence rating of items from Source A to the rating of items from Source B. The ratings indicate here how sure participants are that the item is from Source A.

$$P(R_j|A) = f[P(R_j|B)]. \tag{2}$$

For source recognition, there is another ROC that gives redundant information.

$$P(R_j|B) = f[P(R_j|A)]. \tag{3}$$

With least squares fit analyses of statistics based on Equation 3 generally give the same results as those based on Equation 2. In the figures of the group source ROCs that follow, we present one of the two ROCs and the statistics based on it. When analyzing the statistics from individual ROCs, we make use of the mean statistics from both ROCs.

The standard signal-detection model for item recognition is depicted in Figure 1. The two underlying distributions O (old) and N (new) are arrayed on a discrimination axis (see the top of Figure 1). The overlap of the two distributions produces the discrimination problem for the participant. The distributions are assumed to be normal. The participant is assumed to use the value of the test item on the decision axis and report a number (a confidence rating) related monotonically to that value. The areas to the right of each confidence rating, that is, the areas from the old distribution, $P(R_j \mid \text{old})$, and the new distribution, $P(R_j \mid \text{new})$ are plotted as indicated

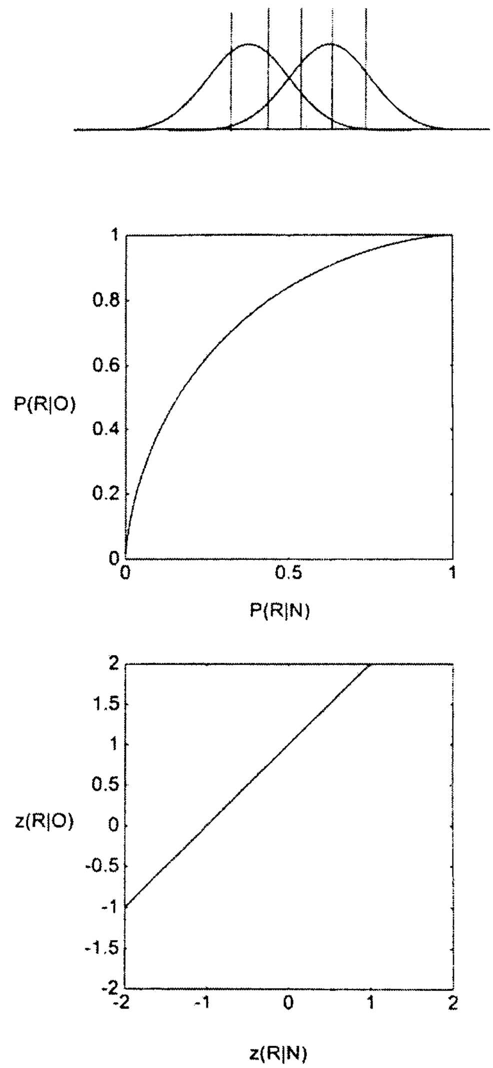

Figure 1. Standard signal-detection model for item recognition. The top panel shows the overlapping new distribution, N, and the old distribution, O. The vertical lines indicate various criteria or confidence levels. The middle panel displays the receiver operating characteristic (ROC) based on the areas to the right of each criterion from each of those distributions (see Equation 1). The bottom panel displays the *z*-ROC, the normalized form of the ROC.

[2] For example, if the ratings go from 1 = *very sure old* to 6 = *very sure new*, old items rated 3 would be computed as the sum of the items rated 3 plus the sum of the items rated 2 plus the sum of the items rated 1, all divided by the total number of old items. A parallel computation holds for new items. The reason for the cumulation is that the area to the right of the criterion set at 3 is assumed to include the areas set to the right of 2 and the area to the right of 1. Examples of the computation are given in Macmillan and Creelman (1991).

by Equation 1. This plot is the ROC. An example of the ROC is shown in the middle of Figure 1. When the values on the axis are converted to the normalized form, as *z*-scores, a *z*-ROC is obtained. An example of the *z*-ROC is shown at the bottom of Figure 1.

Both the ROC and the *z*-ROC give important diagnostic information. If the underlying old and new distributions are normal, the ROC will be convex. When the ROC is replotted in standard score form, as *z*-scores, producing a *z*-ROC, the curvature disappears (assuming the normal distributions of the top panel), and a rectilinear plot appears. If the plot is rectilinear, it indicates that the underlying distributions are normal (or close to normal). The slope of that rectilinear plot is a function of the standard deviations of the underlying distributions. If they are equal, then the slope is 1.00. If the O distribution has a larger standard deviation than the N distribution, then the slope will be less than 1.00. If the O distribution has a smaller standard deviation, then the slope will be greater than 1.00. The *z*-ROC also presents another piece of information. Its intercept is a measure of the accuracy of discrimination (the distance between the O and N distributions). It is labeled d'_2.

Egan (1958) first used the ROC to test models of item recognition. The method was later developed fully by Murdock (1974), Ratcliff et al. (1992), and Ratcliff, McKoon, and Tindall (1994). They made explicit the implications of the model with respect to the ROC. They measured the characteristics of the ROC obtained and determined whether those characteristics conform to the implications of the model. The work of Egan (1958), Murdock (1974), and Ratcliff et al. (1992, 1994) and the work of subsequent investigators (Glanzer, Kim, et al., 1999) have established three regularities for item recognition memory.

1. The ROC for item recognition memory is convex. High threshold theory assumes that old items are either recognized or not on an all-or-none basis. The response to not-recognized old items and to new items is determined by guessing. High threshold theory predicts rectilinear ROCs and is therefore ruled out by convex ROCs (Egan, 1958; Murdock, 1974).

2. When the ROC is replotted with normalized (*z*-transformed) values, producing a *z*-ROC, the *z*-ROC is rectilinear. A rectilinear *z*-ROC is consistent with a simple, normal signal-detection model, that is, a signal-detection model with the standard assumption of underlying normal distributions (Murdock, 1974). A dual-process model for item recognition (Yonelinas, 1999) that included threshold mechanisms predicted concavity of the *z*-ROC. Such concavity would be evidenced by positive quadratic constants when the *z*-ROC is fitted with a quadratic equation. Extensive analysis has found, however, that the predicted curvature does not occur generally (Glanzer, Hilford, Kim, & Adams, 1999; Glanzer, Kim, et al., 1999). Glanzer, Hilford, et al. (1999) reported quadratic constants of the *z*-ROCs of 15 experiments. We have further collected the quadratic constants of the *z*-ROCs from an additional 16 experiments.[3] The mean of the quadratic constants for all 31 experiments is 0.00. The details on the additional quadratic constants are given in Footnote 3.

3. The slope of the item recognition *z*-ROC is less than 1.00 (Ratcliff et al., 1992). Analysis of 41 pairs of slope means from 37 different experiments (see Glanzer, Kim, et al., 1999, Table 3) found that the slope varies, with lower accuracy producing slopes that average 0.80 and higher accuracy producing slopes that average 0.72. We use the value of 0.80 as a benchmark to evaluate the slopes obtained in the following experiments. The usual explanation of slopes less than 1.00 in signal-detection theory is that the underlying distributions, old and new, differ in standard deviation, with the old distribution standard deviation larger than the new.

These findings support a normal signal-detection model as fully adequate for explaining item recognition memory. The model has two main parameters: d' and the ratio of the variance of one underlying distribution relative to the other.

Background on Source Recognition Memory

Interest in source memory was stimulated by the groundbreaking empirical and theoretical work of Johnson and her collaborators on reality monitoring (Ferguson, Hashtroudi, & Johnson, 1992; Hashtroudi, Johnson, & Chrosniak, 1989; Johnson, DeLeonardis, Hashtroudi, & Ferguson, 1995; Johnson, Hashtroudi, & Lindsay, 1993; Johnson & Raye, 1981; Lindsay & Johnson, 1991). Source memory has, moreover, been used in the analysis of effects found in a variety of separate fields. One is the study of brain damage (Janowsky, Shimamura, & Squire, 1989; Mitchell, Hunt, & Schmitt, 1986; Schacter, Harbluk, & McLachlan, 1984; Shimamura & Squire, 1987, 1991); another is the study of aging (Benjamin & Craik, 2001; Brown, Jones, & Davis, 1995; Cohen & Faulkner, 1989; Degli'Innocenti & Bäckman, 1996; Dywan & Jacoby, 1990; Ferguson et al., 1992; Hashtroudi et al., 1989; Henkel, Johnson, & DeLeonardis, 1998; Kausler & Puckett, 1980; McIntyre & Craik, 1987; Schacter, Kaszniak, Kihlstrom, & Valdisseri, 1991). Analyses in those fields have suggested the hypothesis that source and item memory are associated with different brain areas (Glisky, Polster, & Routhieaux, 1995). Source memory has been studied through psychoneurological measurement (Dywan, Segalowitz, & Webster, 1998; Henkel et al., 1998; Janowsky et al., 1989; Johnson, Kounios, & Nolde, 1996; Shimamura & Squire, 1987). It has also been examined in work on child development (Foley, Johnson, & Raye, 1983; Lindsay, Johnson, & Kwon, 1991), the analysis of bilingual performance (Rose, Rose, King, & Perez, 1975; Saegert, Hamayan &

[3] The additional experiments surveyed are the following: Egan (1958), one experiment; Yonelinas (1994), three experiments; Yonelinas, Dobbins, Szymanski, Dhaliwal, and King (1996), two experiments; Yonelinas (1997), three experiments; Yonelinas, Kroll, Dobbins, Lazzara, and Knight (1998), two experiments; Yonelinas (1999), three experiments; and Yonelinas (2001), two experiments. In those cases in which a single experiment yielded two ROCs, the mean quadratic constant was used. We used only the data from normal participants in Yonelinas et al. (1998), from item recognition (not associative recognition) in Yonelinas (1997), and from standard confidence rating procedure (Experiments 1 and 3, not 2a, 2b, 2c) in Yonelinas (2001). The mean quadratic constants are grouped and in the order of the cited studies as follows: 0.03; 0.06, −0.04, and 0.00; 0.06 and 0.06; −0.11, 0.01, and 0.10; 0.03 and −0.05; 0.03, 0.07, and 0.06; −0.07 and 0.03. These results contradict a key prediction of the general dual-process model for item recognition. The model predicts positive quadratics. There are special cases in which item recognition *z*-ROCs show departure from rectilinearity (see Glanzer, Kim, et al., 1999). There are also other tasks-associative recognition (Murdock, 1965; Yonelinas, 1997) and discrimination between single and plural forms of nouns (Rotello, Macmillan, & Van Tassell, 2000) that are distinct from standard item recognition, the detection of generic old items versus generic new items.

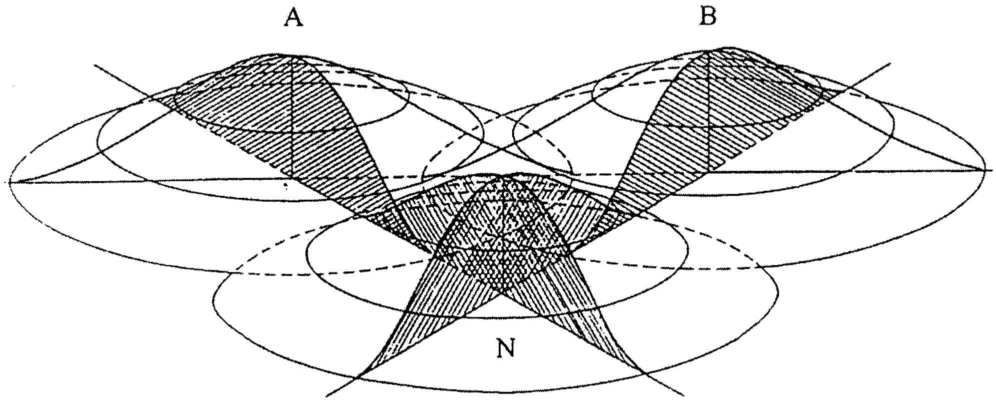

Figure 2. The two-dimensional signal-detection model for both item recognition and source recognition. The means for the new distribution, N, and the two old distributions from Source A and Source B are indicated on the decision axes. From Figure 1 in “Theory of Recognition” by W. P. Tanner, Jr., 1956, *Journal of the Acoustical Society of America, 28,* p. 883. Copyright 1956 by the American Institute of Physics. Adapted with permission.

Ahmar, 1975), and in forensic psychology (Cohen & Faulkner, 1989; Lindsay & Johnson, 1989). The empirical work has been accompanied by theoretical work on several threshold models for source recognition memory. These are discussed later.

A Signal-Detection Model of Source Recognition Memory

Tanner (1956) extended signal-detection theory for sensory discrimination to cover identification. We adapt Tanner’s two-dimensional detection and identification model, obtaining the item and source recognition model schematized in Figure 2. Simple item recognition usually is viewed as involving a single decision axis with two overlapping distributions arrayed (see Figure 1). In Figure 2 these are the noise, or new distribution, labeled N, and the signal, or old distribution, labeled A. When the decision space is expanded to include source recognition, it becomes two-dimensional as in Figure 2 with an additional old distribution, B. There are then three distributions: N, new; A, old from source A; and B, old, from source B. The AN and BN axes are the axes associated with item recognition. Variables that affect item recognition move the means of the A and B distribution on those axes. The AB axis is the axis associated with source recognition. This two-dimensional extension of signal-detection theory to source memory has also been adopted by other investigators (Banks, 2000; DeCarlo, 2000a). We refer to this model as 2D-SDT. The approach falls under the more general theory proposed by Ashby (1992). Banks (2000) used it to explore the geometry of the two-dimensional space.

According to 2D-SDT, in a source recognition test the participant has to decide which of the two old distributions, A or B, to assign a given old item. On the basis of the model and the fact that item recognition *z*-ROCs indicate that old distributions are normal, two predictions follow: (a) The ROCs for source recognition, like the ROCs for item recognition, should be convex; and (b) the *z*-ROCs for source recognition, like the *z*-ROCs for item recognition, should be rectilinear.

Tests parallel to those summarized earlier for item recognition are carried out here for source recognition. The ROC will be plotted. Convexity of the ROC is measured by the quadratic constant obtained from fitting the ROC with a quadratic function. The *z*-ROC is then plotted. It is fitted with both linear and quadratic functions.[4] The linear function furnishes a slope. The quadratic function indicates whether there is any departure from rectilinearity.

Testing the Model: Three Experiments

Preliminary Considerations

We now present three experiments to test 2D-SDT. The experiments furnish critical ROCs and *z*-ROCs. General characteristics of the procedure and analysis of the three experiments are noted first: how the sources were constructed, how the shapes of the ROCs and *z*-ROCs were evaluated, and how the data are presented and analyzed.

Construction of Source Tasks

In the experiments that follow, participants studied items from two sources and were later tested on their ability to remember the sources of studied items. We equated the two sources so that variations in familiarity would not provide a cue to source. To that end, presentation of the two sources was synchronous; that is, both sources were presented in random order in the same study list rather than all of the items from one source being presented before the items from the other source. Within a given experiment, the two sources were also matched on other variables that might affect recognition performance. Specifically, items from the two sources were presented in the same modality (e.g., all auditory), the same

[4] The fitting with linear and quadratic is solely to furnish descriptive statistics, descriptive of the shapes of the ROCs and *z*-ROCs. It is not a theoretical fitting such as that done in the later sections of this article.

number of times (once), and with the same study instructions. The synchronous, matched nature of the two sources is emphasized here because this issue is critical in our later discussion of dual-process theory.

Evaluation of Curvilinearity of the ROC

An ROC for source recognition is defined as $P(R_j \mid A) = f[P(R_j \mid B)]$. Given an ROC, it is possible to determine the extent to which source recognition data conform to a signal-detection model. If the ROC is curvilinear, in particular, if it is convex, it conforms to a normal (i.e., a model that assumes underlying normal distributions) signal-detection model. To evaluate curvature, we fit the ROC with a quadratic equation. For the ROC, the equation is

$$P(R_j|A) = a + b \cdot (R_j|B) + c \cdot [P(R_j|B)]^2. \quad (4)$$

The critical value is the quadratic constant c. If the ROC is convex as predicted by the signal-detection model, c will be negative. If the ROC is rectilinear, as predicted by various threshold models, c will be zero. The sole purpose of this computation is to furnish a descriptive measure of the departure of the ROC from rectilinearity.

Evaluation of Rectilinearity of the *z*-ROC

The second test determines further whether the data conform to a normal signal-detection model. We transform the proportions used to plot the ROCs to *z*-scores and plot a *z*-ROC. If the data conform to a normal signal-detection model, the *z*-ROC should be rectilinear. To test for deviation from rectilinearity, we fit the *z*-ROC with another quadratic equation:

$$z[P(R_j|A)] = a' + b'z[P(R_j|B)] + c' \cdot \{z[P(R_j|B)]\}^2. \quad (5)$$

If the data conform to a normal signal-detection model, the *z*-ROC is rectilinear and the quadratic constant, $c' = 0$. If the *z*-ROC is curvilinear, it will have either a positive quadratic constant c' (if concave) or negative constant c' (if convex).

We also fit the *z*-ROC with linear equation,

$$z[P(R_j|A)] = i + s \cdot z[P(R_j|B)], \quad (6)$$

with i as intercept and s as slope. This furnishes, in the context of a signal-detection model, two additional, informative statistics: the slope and the intercept. The intercept, i, is also referred to as d_2', a measure of accuracy. The slope, s, is the ratio of the standard deviations of the underlying distributions. All fitting will be done by least squares estimation. This method gives results that are essentially the same as those for maximum likelihood estimation.[5]

Data Presentation

For each of the experiments that follow, the confidence rating data are presented in two forms. One form, group, is based on pooling the responses of all the participants, summing their responses across each of the confidence levels and generating a single, group ROC. Corresponding to each group ROC is a group *z*-ROC based on the same set of pooled responses. These ROCs give a simple overall picture of the data. For statistical analysis, however, we fit each participant's responses separately, generating an ROC and a *z*-ROC for that participant. Each ROC and *z*-ROC is individually fitted, the statistics describing each are derived, and the means of those statistics are presented and analyzed. For example, an index of curvature, the quadratic constant, is computed for each participant's ROC and the mean of those indices used to describe the set of ROCs. The standard errors of those indices are used to test the means. As will be seen, the statistics from the group ROCs and the individual ROCs give concordant information.

General Procedure and Analysis

We used two different sensory modalities to present stimuli. In Experiments 1 and 2 the stimuli were auditory—words spoken either by a man or a woman. In Experiment 3 the stimuli were visual—words presented at either the top or the bottom of a computer screen. Two different testing procedures were also used. In Experiment 1 we tested both item and source recognition in sequence: An item recognition test of each word was followed by a source recognition test of that word. Test lists contained both old and new words, and during test, participants indicated which words they thought were old and which were new. For each item selected as old, participants then indicated the source of the word using a confidence rating scale. In Experiments 2 and 3 we tested source recognition without a preceding item recognition test. Following the study list, the participants were given a list that contained only old (studied) words, and they were asked to indicate the source of each using a confidence rating scale.

Once the ROCs are obtained, analysis consists of two stages. First, the shape of each ROC is examined. Second, each ROC is transformed to normal form, referred to as a *z*-ROC, and its shape is examined.

Experiment 1: Item and Source Recognition, Auditory

In Experiment 1 participants heard a list of words, in which each of the words on the list was spoken by either a man or a woman. This variation in source has been found to be effective as a basis for source judgments (Geiselman & Bellezza, 1976). After hearing the words, participants were tested successively on item recognition and source recognition.

Method

Participants heard a list of 180 study words, 90 in a man's voice and 90 in a woman's. During the test that followed, participants viewed a random, mixed list of 180 old and 180 new words on a monitor and were instructed

[5] According to Glanzer, Kim, et al. (1999):

> Maximum likelihood estimates . . . have an advantage in that they assume, as is the case, that both axes of the ROC are subject to error. Least squares estimates assume that the *x*-axis is error free. We applied both estimation procedures to the data. The results for both estimates were in very close agreement. The correlation between the 10 maximum likelihood intercepts and the 10 least squares intercepts obtained in the four experiments described next was .944; the correlation between the 10 slopes was .994. Statistical analyses of the intercepts and slopes obtained from the least squares estimates gave the same results as those based on the maximum likelihood estimates. (p. 501)

to identify which were old and which were new (yes/no item recognition). If they identified a word as old, they then indicated whether it had been spoken by the man or the woman, using a 6-point confidence rating scale.

Materials

Three hundred and sixty words were digitally recorded into computer memory by both a man and a woman (720 words in total) and were used as the main list for the experiment. Forty-five additional words were recorded in the same way for practice and filler items. Each word was stored as an individual sound file. The words were chosen from Paivio, Yuille, and Madigan (1968). The words had an average word length of 6.42 letters, average frequency of 43.47 (words per 10,000), and average concreteness rating of 5.02.

Procedure

The experiment began with a 5-word practice study list and a 10-word test list to familiarize the participants with the procedure. The practice study and test procedure was the same as that for the main list.

For each participant, 180 words were selected at random from the full list of 360 words, randomly ordered, and presented in the study list. The voicing of the study list items was also selected at random for each word with the restriction that half were in the male voice, half in the female voice. The study list was presented through headphones connected to the computer. Participants were told to listen to the words and that a test would follow. While the words were presented through the headphones, the computer screen was blank. A 1,000-ms interstimulus silent interval separated successive words.

The recognition test, presented visually on the computer monitor, followed the study list directly. Each test list consisted of the 180 study words plus 180 new words (the remainder of the main list), all mixed and randomly ordered. All of the items appeared in the middle of the screen in uppercase letters. Participants viewed the words one at a time and indicated whether each word was old or new, respectively, by pressing a key marked "O" or a key marked "N." If the participant responded "O," a second display appeared asking whether the item had been spoken by the man or the woman. Participants selected keys on the keyboard marked "1" through "6": 1 = *very sure male*; 2 = *moderately sure male*; 3 = *a little sure male*; 4 = *a little sure female*; 5 = *moderately sure female*; and 6 = *very sure female*. Each of the choices appeared in a row in the center of the screen. Below each number was the criterion's description as in the above sentence. Participants in this experiment and the other experiments were instructed to use all six confidence ratings. The test was self-paced. We used a simple yes/no test for item recognition (no confidence rating) to avoid influence of item confidence ratings on source confidence ratings.

Study and test lists, as noted earlier, were randomized individually for each participant. The study list began with five untested filler items and ended with five untested filler items. (By error, four of these words here, and three in Experiment 2, also appeared in the pool of words used to generate the test lists. These words were excluded from the analyses, reducing the test list by up to four words.)

Participants

Responses were obtained from 17 students from an undergraduate psychology class who participated to satisfy a class requirement. Participants had been speaking English since the age of 10 years or earlier. This description of the source of the participants and their language background holds for the participants in all the following experiments. The data set of 1 additional participant who did not use the rating scale as instructed was not used.

Results

The rating responses summed across all participants are presented for this and the following experiments in the Appendix. ROCs based on the group data are presented in Figure 3. The top left panel gives a group ROC obtained by summing across the responses of all the participants and plotting a single ROC[6] − $P(R_j|\text{male}) = f[P(R_j|\text{female})]$. The top right panel shows the corresponding group *z*-ROC. The group ROC is convex: The quadratic constant is negative, as indicated in the figure. Although Experiment 1's group ROC in Figure 3 is clearly convex, it is slightly flattened. The flattening of the group ROC corresponds to the concavity found in the *z*-ROCs for Experiment 1. The quadratic constant for the *z*-ROC is positive.

The findings on the shapes of the group ROC and *z*-ROC are supported by the statistics obtained from the individual participants' ROCs and *z*-ROCs. The statistics are given on the first line in Table 1. In this and the experiments that follow, these statistics were obtained and tested in the following way. The ROC for each participant was computed. Each of the statistics from the ROC and the *z*-ROC was then tested as described next. Statistical significance was set at the .05 level.

As noted earlier, in the case of source recognition, two ROCs $P(R_j|\text{male}) = f[P(R_j|\text{female})]$ and $f[P(R_j|\text{female}) = f[P(R_j|\text{male})]$ can be plotted and statistics derived from them. We plotted both for each of the experiments that follow and tested the statistics obtained for differences between the two functions. In all cases but two, in Experiment 3, there was no marked or statistically significant difference in the statistics obtained from the two functions. For Experiment 3 the means of the statistics for the two functions are analyzed and reported separately.

To evaluate the extent of convexity of the ROCs, we fitted each individual ROC with a quadratic equation. These equations gave 17 pairs of quadratic constants with a mean of −1.26. Its difference from zero is statistically significant, $t(16) = 5.77$.

The *z*-ROC for each individual participant was also computed and each fitted with a quadratic equation. The mean quadratic constant, 0.24, differs significantly from zero, $t(16) = 6.02$. This concavity is replicated in the experiments that follow. This is the first difference from the regularities found for item recognition. It is important to note that although the departure from rectilinearity in the *z*-ROC will be found to be general, it is a small departure. This can be seen by measuring the amount of variance accounted for by that quadratic component. In these data, a rectilinear fit accounts for 98.1% of the variance in the top right panel of Figure 3. A quadratic fit increases the amount of variance accounted for to 99.99% of the variance. In other words, the quadratic deflection accounts for only 1.9% of the variance in the *z*-ROC. This type of analysis of the group data was carried out for Experiments 2 and 3. Both indicate that the quadratic component accounts for a small amount of additional variance, 2.8% in Experiment 2, 0.8% in Experiment 3. The departure from rectilinearity, although small, is, however, reliable and a revision of the 2D-SDT model is needed. This revision is presented later.

The individual *z*-ROCs were also fitted with linear equations. The linear fits gave a slope of 0.99. This finding indicates a second difference between source recognition and item recognition. A large number of studies have shown that the slope of the *z*-ROC for item recognition averages 0.80 or less (Glanzer, Kim, et al., 1999;

[6] The other group ROC $P(R_j \mid \text{female}) = f[P(R_j \mid \text{male})]$ gives, as noted earlier, redundant information.

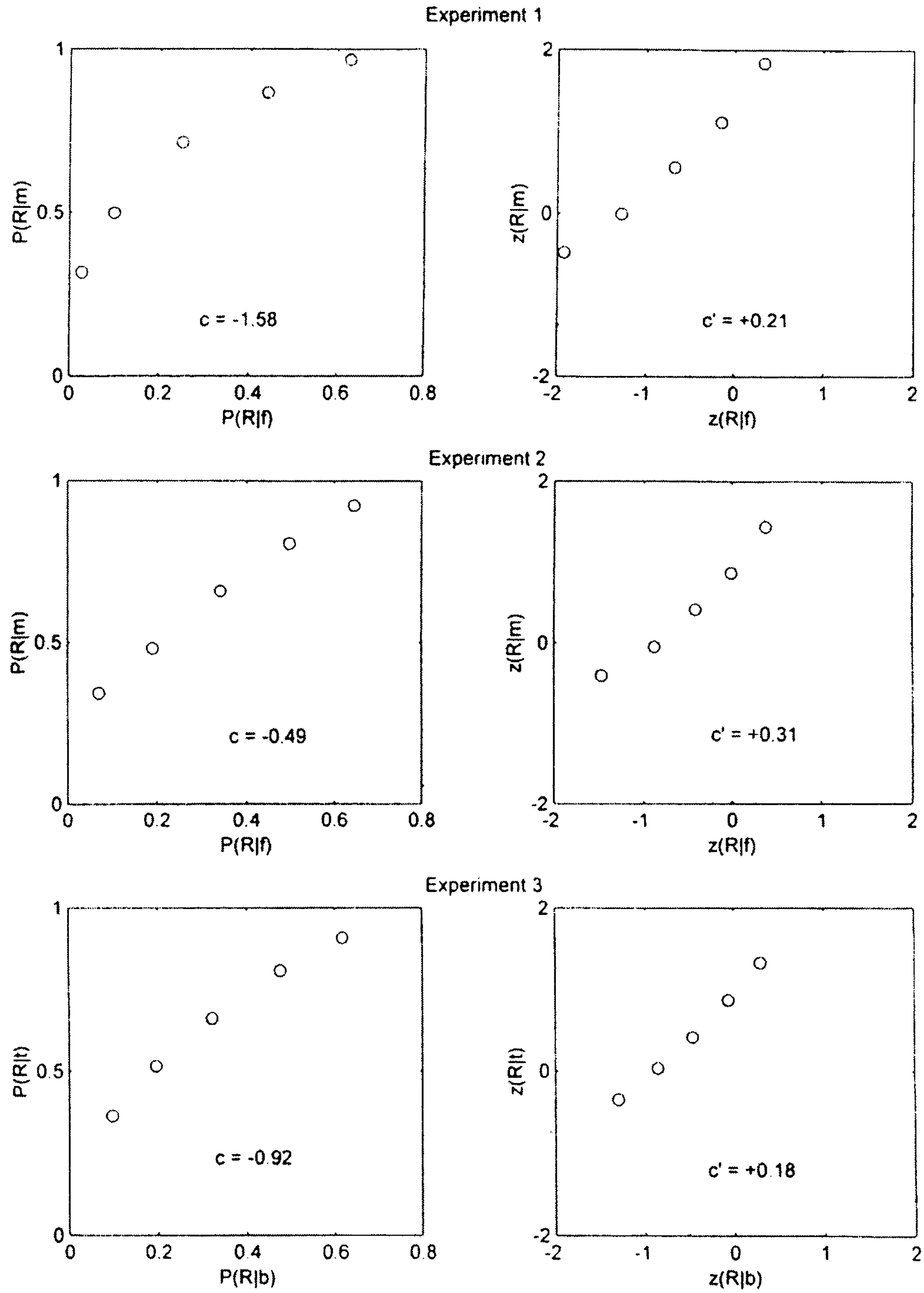

Figure 3. Group receiver operating characteristics (ROCs) and *z*-ROCs of Experiments 1, 2, and 3. Quadratic constants for the ROCs are indicated by c, for the *z*-ROCs by c′.

Ratcliff et al., 1992, 1994). The difference between the obtained mean slope of 0.99 and the benchmark item recognition slope of 0.80 is statistically significant, $t(16) = 4.65$. The mean slope, moreover, does not differ significantly from 1.00, $t(16) < 1.00$. The statements about slope are limited somewhat by the departure of the *z*-ROC from rectilinearity. The departure is, however, slight, as indicated above, and the slope value is informative. It indicates that the underlying distributions are likely to be equal in variance.

The difference in slopes between item recognition and source recognition makes good sense. In signal-detection theory, the slope of the *z*-ROC equals the ratio of the standard deviations of the two underlying distributions. In item recognition, the old distribution's

Table 1
Statistics for the ROCs and z-ROCs of Three Experiments

Experiment	ROC Quadratic constant	*z*-ROC Linear slope	*z*-ROC Quadratic constant
1	−1.26 (0.22)	0.99 (0.04)	+0.24 (0.04)
2	−1.21 (0.32)	1.01 (0.03)	+0.29 (0.06)
3a[a]	−1.38 (0.33)	1.08 (0.04)	+0.21 (0.09)
3b	−1.06 (0.24)	0.92 (0.03)	+0.16 (0.06)

Note. Standard errors are in parentheses. ROC = receiver operating characteristic.
[a] The statistics for both ROCs, $P(R_j|\text{top}) = f[P(R_j|\text{bottom})]$ and $P(R_j|\text{bottom}) = f[P(R_j|\text{top})]$ are reported because the two ROC quadratic constants and the two z-ROC linear slopes differ at a statistically significant level. (See text).

variance may be presumed to be increased by the variance associated with learning during the study trial. Therefore, $\sigma(\text{new})/\sigma(\text{old}) < 1.00$. In the case of source recognition, the relation between the underlying distributions according to 2D-SDT is quite different. In source recognition we have two old distributions that have the same initial composition (randomly selected items) and the same history. Both have an increment in variance produced by learning. Therefore, they should have the same variances with $\sigma(A)/\sigma(B) = 1.00$. The slope of the *z*-ROC should therefore be 1.00. This finding would be expected for cases in which the two sources are synchronous, matched and in the same sensory modality (e.g., auditory), as in this experiment. If they were not, as in the case of sources that are asynchronous, or unmatched, or in different modalities, slopes of 1.00 would not be expected. We later comment on two such cases reported by Yonelinas (1999).

In our experiments, the two sources were matched on variables likely to affect recognition memory, to eliminate the possibility that differences between them in “familiarity” or “strength” would provide cues to source. We checked on such differences using the item recognition data. Each participant furnished two d's: one for male-voiced old items versus the new items on the preliminary item recognition test, the other for the female-voiced old items. The two means are 1.04 and 0.99. The difference between the two is not statistically significant, $F(1, 16) = 0.72$, $MSE = 0.03$. There is no evidence of any familiarity or strength difference.

Before considering theoretical interpretation, we report results from two more experiments. The critical pattern of results found here is replicated in the experiments that follow.

Experiment 2: Source Recognition Without Item Recognition, Auditory

In this experiment we simplified the procedure by testing source recognition directly, uncoupling it from item recognition. The participant did not do item recognition first on each test item as in Experiment 1 but was given only old items, defined as such, and responded only as to the source of each item. For the 2D-SDT model, this means that the experiment focuses only on decision axis AB in Figure 2. The ROC is still expected to be convex. This simplification is important for two reasons. One is that it tests the generality of our findings across paradigms. The other is that it simplifies the testing of theories of source recognition.

Method

Two 90-word lists were presented for study and then were tested for source recognition. Five filler items appeared at the beginning and end of each of the study lists and at the beginning of each of the source recognition test lists. A short item recognition test was also given. The main purpose of this test was to prevent participants from attending to only one of the two voices in the study lists for source recognition. Pilot work indicated that the participants developed this efficient strategy, effectively halving their study list when they knew that item recognition would not be required.

Materials

The words used were the same as those used in Experiment 1.

Procedure

The experiment began with a 7-word practice study list followed by source and item test lists to familiarize participants with the procedure. Each participant was given two study/test blocks. Each block consisted of a study session, a source recognition test, and then a short item recognition test. The study lists were constructed in the same way as those in Experiment 1 except that only 90 words were presented in each study list and instead of a 1,000-ms interstimulus interval, a 1,250-ms interstimulus interval separated successive words. During the interstimulus interval, a row of asterisks appeared in the center of the computer screen.

After the study words were presented, participants were given a source recognition test. In this test all of the items were old—words that appeared in the study list in a rerandomized order. The participants knew that all the items were old, from the instructions and from having had a preliminary practice list. The words were rerandomized so that the order of words in the study list and in the test list was different. For each word, participants were instructed to determine whether the item had been spoken by the man or the woman. Participants indicated their choice using a 6-point confidence rating scale as in Experiment 1.

Following the source recognition test, participants were given a brief self-paced yes/no item recognition test. The item recognition test consisted of 20 old words and 20 new words in random sequence.

The study and test lists were randomized separately for each participant. The old/new item recognition test lists were created by randomly selecting 15 words from the study list and adding to them an equal number of new words. Instructions appeared on the screen before each study list and each test list to remind participants of their tasks for each upcoming section.

Participants

Forty-four undergraduate students participated.

Results

The group ROC and *z*-ROC for this experiment are shown in the left and right middle panels of Figure 3. We collapsed the data for the two blocks. Preliminary analysis of the individual participants’ data showed that blocks had no effect on any of the ROC statistics. The picture is similar to that for Experiment 1. The group ROC is again convex (quadratic constant negative), and the group *z*-ROC is again slightly concave (quadratic constant positive).

The analysis of individual participants’ ROCs was the same as that in Experiment 1. Each participant’s ROCs were fitted with quadratic equations to obtain 44 pairs of quadratic constants. The corresponding individual *z*-ROCs were fitted with linear equations to obtain slopes and quadratic equations to obtain quadratic constants. The second row of Table 1 summarizes the means and standard errors obtained from the individual ROCs and *z*-ROCs. The mean quadratic constant based on the individual ROCs was −1.21. This differs significantly from zero, $t(43) = 3.81$. The ROC is convex.

The quadratic fitting of the individual *z*-ROCs gave a mean quadratic constant of 0.29. This again differs significantly from zero, $t(43) = 5.07$. The linear fitting of the *z*-ROCs gave a mean slope of 1.01, which does not differ at a statistically significant level from 1.00 but does differ from the benchmark item recognition slope of 0.80, $t(43) = 7.95$.

We checked the data from the item recognition test that followed the source recognition test. The mean d' for male-voiced old versus new items was 2.55; for female-voiced old versus new items, 2.38. The difference between these means was, again, not statistically significant, $F(1, 43) = 1.74$, $MSE = 0.36$. There is again no evidence of a difference in familiarity or strength of the two sources.

Experiment 2 replicates, with a different procedure, the findings of Experiment 1: a convex ROC and a concave *z*-ROC with a slope near 1. The change in procedure had no effect on the overall pattern of results.

We noted earlier that there has been extended discussion of signal-detection models for identification. These discussions have pointed out complexities in such models. Batchelder, Riefer, and Hu (1994) argued against the use of signal-detection theory for source recognition. They claimed that a theory involving a two-dimensional space such as the one depicted in Figure 1 presents major analytic problems. Experiment 1 involves a two-dimensional space. The participant carries out both item recognition and source recognition on the same items. The present experiment isolates the source detection task in a one-dimensional space, like the one-dimensional space assumed for item recognition. The participant knows that all items in the source recognition test are old. Although a subsequent item recognition test is given, the source recognition is made independent of the item recognition. The results for the two-dimensional decision space of Experiment 1 and the one-dimensional space of Experiment 2, however, give the same key results. Measures of the form of the ROC and form of the *z*-ROC show the same regularities.

Experiment 3: Source Recognition Without Item Recognition, Visual

In this experiment we extended the findings of Experiment 2 using a different sense modality, with visual instead of auditory study items. We produced two different sources by presenting study list words in different positions on the computer monitor—top or bottom of the screen. In the source recognition test, the participants were instructed to indicate the study list word's position. Except for stimulus modality, the procedure was essentially the same as that of Experiment 2. After the source recognition test was completed, it was followed by a token item recognition test.

Method

Two hundred and eighty words were drawn, and the set was randomly divided in half for use in each of the two study/test blocks. The two study lists each consisted of 140 words. The appearance and position of the study words were randomized independently for each participant. The first 10 and last 10 study list words were omitted from the test to eliminate primacy and recency effects. The remaining 120 words were rerandomized and then presented as the test list.

Materials

Two hundred and eighty words were selected from Paivio et al. (1968). The words had an average word length of 6.25 letters, average frequency of 43.80 (words per 10,000), and average concreteness rating of 5.02.

Procedure

The procedure, except for stimulus modality, number of items, and number of buffer items, was the same as that of Experiment 2. The session started with a practice study/test block using the same procedure as the main study/test block, but with only 20 words in the study list. During each of the two study blocks that followed, participants viewed a list of 140 words. The words were presented one at a time in one of two locations: the top of the monitor or the bottom of the monitor, 70 in each position. The words appeared in lowercase letters for 3,000 ms with a 250-ms interstimulus interval.

A source recognition test with a 6-point confidence rating scale followed the study list. The test list consisted of only old items, the words from the study list. Each test item appeared in the center of the screen in lowercase letters. Participants were instructed to decide for each word whether it had appeared at the top or bottom of the monitor during study. On the confidence rating scale, 1 = *very sure on the top* and 6 = *very sure on the bottom*. At the end of the test, the participant's score was presented on the screen. As in Experiment 2, after the source recognition test, a short yes/no item recognition test was presented. As noted earlier the main purpose of this test was to prevent participants from attending to only one source.

Participants

Thirty undergraduate students furnished data for analysis. The data of 2 participants were not included in the analysis. These 2 participants used only two positions on the rating scale. Their restricted ratings did not permit the plotting of ROCs for them.

Results

As in Experiment 2, the data for the two source recognition study/test blocks were combined after analysis showed that presentation order had no effect on any of the ROC statistics. The plots of the group ROC and *z*-ROC are shown in the lower panels of Figure 3. The plots are similar to those found in the preceding two experiments. The ROCs are convex (quadratic constant negative). The *z*-ROCs are concave (quadratic constant positive).

The means of the statistics for the individual ROCs and *z*-ROCs (shown in rows 3a and 3b of Table 1) are similar to those of the preceding experiments. The analysis of the statistics is, however, more complex. As noted earlier, two ROCs can be plotted for each set of data. Here the two are $P(R_j|\text{top}) = f[P(R_j|\text{bottom})]$ and $P(R_j|\text{bottom}) = f[P(R_j|\text{top})]$. The means in Table 1 are reported separately for the two functions because on two measures—the ROC quadratic constant and the *z*-ROC linear slope—the mean statistics for the two functions differ at a statistically significant level. However, on almost all of the critical tests, they furnish the same answers. Both ROC quadratic constants are significantly lower than zero, $t(29) = 4.20$ and $t(29) = 4.34$. Both *z*-ROC quadratic constants are significantly higher than zero, $t(29) = 2.29$

and $t(29) = 2.59$. The z-ROC linear slopes are both significantly different from the benchmark item recognition slope of 0.80, $t(29) = 7.77$ and $t(29) = 4.05$. The only departure is that one slope is significantly higher than 1.00, $t(29) = 2.17$, and the other lower than 1.00, $t(29) = 2.53$.

As in the preceding experiments, a check for a difference in the item recognition accuracy for the two sources found none. The mean d' for the top and for the bottom was the same, 0.61, and the $F(1, 29) < 1.00$. No difference in familiarity of items from the two sources is evident.

The results replicate the main findings of the preceding experiments. The change in the modality of the source has not affected the key findings. The three regularities hold. The ROC is convex. The z-ROC is concave. The linear slope of the z-ROC is higher than the average slope found for item recognition and close to 1.00.

Other Data on the Regularities

There are other confidence rating data that concur with our findings of the three regularities, those of Yonelinas (1999), Qin, Raye, Johnson, and Mitchell (2001), and Slotnick, Klein, Dodson, and Shimamura (2000). Statistics based on those data are presented in Table 2. Yonelinas (1999) reported the results of four source recognition experiments. Experiment 1 is most similar to our experiments. The sources (left and right on the monitor) were synchronous (both presented in one list) and matched.[7] Experiment 2 was asynchronous (two successive lists, one spoken by a man, the other by a woman—a list differentiation structure), but the asynchrony did not appear to have any effect here, and the list items can be considered matched.[8] Experiments 3 and 4 were, however, different from the other experiments in Table 2. Experiment 3 was asynchronous, two successive lists as in Experiment 2, but here the two lists were made to differ markedly in "strength." The first list was presented twice, the second once. Experiment 4 was asynchronous (two successive lists) presented 5 days apart. This meant that the two lists were made to differ markedly in strength because the first list was tested after a 5-day delay. The effect of this difference in treatment is discussed shortly.

Table 2
Statistics for the ROCs and z-ROCs for the Experiments of Mather et al. (1999), Qin et al. (2001), Slotnick et al. (2000), Yonelinas (1999), and the Present Study

Experiment	ROC: Quadratic constant	z-ROC Linear slope	z-ROC Quadratic constant
Mather et al. (1999)[a]			
1	−2.04	0.89	+0.04
2	−5.42	0.98	−0.40
Qin et al. (2001)			
1	−2.58	1.18	+0.07
2	−2.60	1.14	+0.03
Slotnick et al. (2000)			
1	−0.52	0.97	+0.09
2	−1.56	0.94	+0.08
3	−0.84	0.98	+0.25
Yonelinas (1999)			
1	−0.62	0.98	+0.41
2	−0.45	1.03	+0.26
3	−0.25	0.75[b]	+0.22
4	−0.47	0.70[b]	+0.15
Present			
1	−1.26	0.99	+0.24
2	−1.21	1.01	+0.29
3[c]	−1.22	1.00	+0.18

Note. The present study statistics are means from individual receiver operating characteristics (ROCs) and z-ROCs. All the others are based on group ROCs and z-ROCs. The Slotnick et al. statistics were presented in that article. The Mather et al., Qin et al., and Yonelinas statistics were computed by us on the basis of their ROC and z-ROC plots.
[a] Mather et al. (1999) data were analyzed and presented in Qin et al. (2001). [b] The reasons for the large departure of these slopes from 1.00 is discussed in the text. [c] These are the means of the values reported in Table 1.

Slotnick et al. (2000) reported three experiments in which the two sources were male and female voices. The sources were synchronous and matched, with the item recognition accuracy for the two sources equivalent. The participants carried out both old–new confidence ratings and source confidence ratings for each test word. The relevant mean statistics for the group source ROCs were collapsed over item ratings.

Qin et al. (2001) reported four sets of ROC data. Two of these were obtained from an experiment by Mather, Johnson, and DeLeonardis (1999), and two were from their own experiment. In Mather et al.'s experiment, the two sources were two women reading statements on a videotape. The two ROCs were generated by differences in the instructions to the participants (as to whether they focused on their own feelings or the speakers' feelings). In Qin et al.'s experiment, the two sources were also two women reading statements on a videotape. There were two groups in that study: one used a standard confidence rating scale, whereas the other used a rating of amount of information about their memory of the speaker.

All of the statistics in Table 2 except for the present study are based on group data, that is, the ratings summed across all of the participants to give a single ROC and z-ROC. The present study's statistics are based on means of individual ROCs.

All 14 of the ROC quadratic constants are negative, that is, all 14 indicate that the ROCs are convex. Thirteen of the 14 z-ROC quadratic constants are positive, indicating concavity. The probability of 13 out of 14 positive results by chance is, using a binomial test, 0.0009. The linear slopes show some variability. The major variants are found in Yonelinas's Experiments 3 and 4, with slopes of 0.75 and 0.70, well below 1.00. These are similar to the slopes found in item recognition, and that similarity makes good sense in the context of signal-detection theory. As noted earlier, the fact that item recognition produces z-ROC slopes of 0.80 or less can be explained as resulting from the additional operation imposed on the old distribution, namely, study. This increases the variance of that distribution and therefore reduces the slope. In Yonelinas's

[7] Evidence of the equivalence of the two sources, that is, the absence of strength or familiarity differences, can be seen in the item recognition accuracy measures, computed on the basis of group z-ROCs. The item recognition d'_2 for the left items is 1.28, for the right items is 1.25.

[8] As in Yonelinas's (1999) Experiment 1, the item recognition accuracy measures do not indicate any difference in strength or familiarity. The item recognition d'_2 for the first list was 0.94, for the second list 0.91.

Experiment 3, one source's items were studied twice, the other studied once. In his Experiment 4, one source's items were subjected to 5 days' delay and the other source's items were tested immediately after study. In both cases, the two sources are not matched. An additional operation imposed on one of the two lists changes the variation of one underlying distribution. The mean of the remaining 12 slope means is 1.01, and the generalization holds: ROCs are convex, *z*-ROCs are concave, and *z*-ROC linear slopes are near 1.0.

Revision of the Signal-Detection Model and Test of the Revision

Although the 2D-SDT model covers 98% to 99% of the variance in the ROCs, the concave *z*-ROCs indicate that the model requires revision. The revision has to meet three criteria. It must produce (a) convex ROCs, (b) concave *z*-ROCs, and (c) *z*-ROC linear slopes at or near 1.00.

We first attempted to revise the model by replacing the normal distribution with other distributions—logistic, poisson, chi-square, and binomial (Egan, 1975). None of these satisfy criterion b. None produce concave *z*-ROCs. An extreme value distribution gives concave *z*-ROCs but does not give *z*-ROC slopes of 1.00 (DeCarlo, 1998). There is, however, a revision of 2D-SDT that produces distributions that satisfy the three criteria and that fits well in the general framework of signal-detection theory: a mixture model.

Revised Model

The motivation for the revised model comes from considering the role of attention in source memory: "stress or divided attention may disrupt normal perceptual and reflective processes, resulting in relatively impoverished information encoded, information from which source could later be derived" (Johnson et al., 1993, p. 5). Johnson et al. continued with a description of a case in which a stimulus arrives lacking source information, for example, a statement overheard without sight of the speaker at a cocktail party (see also Zaragoza & Lane, 1991). Both of these effects—inattention and absence of attendant source information—are captured in an extension of signal-detection theory, a mixture model proposed by DeCarlo (2000b, 2002) for a range of applications to recognition memory. For source recognition, the model assumes that the sources of some items are attended to during study, whereas some are not. Alternatively, some items come with associated source information, some do not. Items of the first class from distributions A and B are normally distributed with means in opposite directions away from zero on the AB axis. Items of the second class, N′, are normally distributed, with a mean of zero.[9] The AB axis of Figure 1 therefore has on it another distribution, N′, placed midway between A and B. This arrangement is shown in Figure 4.

The experimenter analyzing response data for source recognition is dealing with two mixture distributions: one a mixture of A and N′ and the other a mixture of B and N′. These mixture distributions are nonnormal. DeCarlo (2000b) showed that such distributions can generate ROCs that are convex and *z*-ROCs that are concave with linear slopes of 1.00. There is only one new parameter in this model: the proportion of effective or attended to items, λ. The equations that follow from this formulation are

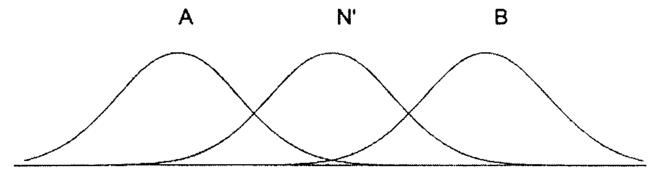

Figure 4. Revised (mixture) 2D-SDT (two-dimensional signal-detection theory) model for source recognition. For the sake of clarity, the representation has been reduced to one dimension with the N distribution and the AN and BN axes eliminated. The distribution of unattended items, N′, has been projected on the AB axis.

$$P(R_j|A) = \lambda \int_{-\infty}^{c_j} G(x|A)dx + (1-\lambda) \int_{-\infty}^{c_j} G(x|N')dx, \quad (7)$$

$$P(R_j|B) = \lambda \int_{-\infty}^{c_j} G(x|B)dx + (1-\lambda) \int_{-\infty}^{c_j} G(x|N')dx, \quad (8)$$

where R_j is a rating from confidence level j to the highest rating; c_j is the criterion at rating j; G is a normal, or Gaussian, distribution; λ is the proportion of effective or attended to items, $0 \le \lambda \le 1$; A is Source A, B is Source B, and N′ is a set of items that do not bear source information. This is a simple form of a range of mixture models presented by DeCarlo (2002). It will suffice for the purposes of this presentation.

Maximum likelihood fits of the model (Matlab, 1995; Vermunt, 1997) to the data of the present experiments give the results summarized in Table 3 and Figure 5. None of the likelihood ratio goodness-of-fit tests, G^2 in Table 3, are statistically significant. The fit accounts for 99.6%, 99.7%, and 99.99% of the variance in Experiments 1, 2, and 3, respectively. Also included in Table 3 are maximum likelihood fits of the unrevised 2D-SDT model. All the fits indicate a statistically significant deviation of the data from the fitted points. Moreover, the models are nested, with $\lambda = 0$ for the unrevised 2D-SDT model. Therefore, the G^2s for the mixture model can be subtracted from the G^2s for the unrevised model. Each of these difference G^2s has one degree of freedom. As indicated in the table, each difference is statistically significant. This indicates a statistically significant improvement of fit with the revised model. Details on this evaluation procedure may be found in Bishop, Fienberg, and Holland (1975).

Alternative Models

There are published formal models of source recognition memory that involve threshold mechanisms. These, the family of multinomial processing models and the dual-process model, all reduce to simple threshold models, as will be shown. Threshold models are based on the assumption that the states underlying recognition decisions are few and discrete. A test item is, for example, assumed to produce one of two states—either recognized or not. This assumption distinguishes threshold models from signal-

[9] This added distribution does not imply that the model is a threshold model. Threshold models involve discrete, noncontinuous states. This model assumes continuity as does the standard signal detection model.

Table 3
Comparison of Mixture Model and Simple Normal Model Parameters, and Evaluation of Fits of ROCs for Experiments 1, 2, and 3

Model and experiment	d_{AB}	λ	*df*	G^2
Revised model				
1	3.9	0.47	3	2.00
2	4.8	0.32	3	6.82
3	3.5	0.37	3	3.24
Unrevised model				
1	1.4	0.00	4	12.75*
2	0.9	0.00	4	128.28*
3	0.9	0.00	4	25.78*
Difference of G^2				
1	—	—	1	10.75*
2	—	—	1	121.46*
3	—	—	1	22.54*

Note. G^2 = likelihood ratio goodness-of-fit test. ROC = receiver operating characteristics.
* $p < .05$.

detection models. Signal-detection models have in common the contrasting assumption that the states underlying recognition decisions are many and, in most cases, continuous. A test item is judged to be in any one of a range of states, each associated with a different measure, such as strength. Continuity of underlying states produces a continuous decision axis. The difference in assumptions leads to markedly different predictions about the shape of ROCs.

The contrast between models based on threshold theory and those based on signal-detection theory is covered in an extensive literature on sensation and memory. That literature has been summarized by Green and Swets (1974) and more recently by Macmillan and Creelman (1991). The first work contrasting these two types of theory for recognition memory was by Egan (1958). Egan demonstrated that the ROC obtained for item recognition was curvilinear, as predicted by a signal-detection model, and not rectilinear, as predicted by a threshold model. As a result of that and subsequent work, signal-detection theory has become the generally accepted basis for explaining item recognition. (See its use in the theories of Gillund & Shiffrin, 1984; Glanzer, Adams, Iverson, & Kim, 1993; Hintzman, 1988; Murdock, 1982.) Threshold theory mechanisms have, however, been resurrected as appropriate for source recognition memory, for example, the multinomial processing models of Batchelder and Riefer (1990) and Batchelder, Hu, and Riefer (1994). Besides the models just mentioned, there are several other threshold models discussed in the literature (Green & Swets, 1974; Krantz, 1969; Macmillan & Creelman, 1991; Murdock, 1974). We consider here, however, only those that are relevant to source recognition.

The use of confidence ratings to evaluate threshold models for source recognition follows the procedure adopted by Egan (1958) to evaluate a threshold model of item recognition. It is assumed that in a threshold framework the participants vary their guessing responses, assigning them to different rating levels.

A Threshold Model, Derived

There is no free-standing, simple threshold model to cover confidence ratings in source recognition.[10] Batchelder and Riefer (1990), however, presented a family of multinomial processing models based on high-threshold mechanisms for item and source recognition jointly. Models from this family have been widely applied. They have been fitted to source recognition data by Batchelder and Riefer (1990), Bayen, Murnane, and Erdfelder (1996), Johnson, Kounios, and Reeder (1994), and Riefer, Hu, and Batchelder (1994). The models are set up to handle yes/no data. We can, however, obtain information relevant to that family of models and other threshold models by carrying out two steps.[11]

First, we simplify the theory to expose its implications for source recognition alone. Second, we extend the simplified theory to cover confidence rating. The simplification is done by considering the case in which all items are known by the participant to be old. This is the condition that was set up in Experiments 2 and 3. With this simplification the following equations hold.

$$P(\text{“A”}|A) = d_1 + (1 - d_1)a, \tag{9}$$

$$P(\text{“B”}|B) = d_2 + (1 - d_2)(1 - a), \tag{10}$$

where P(“A” | A) is the probability of saying “A” to a Source A item; P(“B” | B) is the probability of saying “B” to a Source B item; d_1 is the detection of a Source A item; d_2 the detection of a Source B item; and a is the bias toward saying “A.”

Extension of these equations to confidence ratings gives the following:

$$P(R_j|A) = d_1 + (1 - d_1)a_j, \tag{11}$$

$$P(R_j|B) = d_2 + (1 - d_2)a_j, \tag{12}$$

where the R_j is a rating from confidence j to the highest confidence rating and a_j is the probability of choosing a rating from confidence level j to the highest rating in the absence of source detection.

These equations give rise to the following rectilinear relation:

$$P(R_j|A) = \left(\frac{d_1 - d_1d_2 - d_2}{1 - d_2}\right) + \left(\frac{1 - d_1}{1 - d_2}\right)P(R_j|B). \tag{13}$$

This implies a rectilinear ROC (not *z*-ROC). All experiments in Table 2 have negative quadratic constants, which indicate convex, not rectilinear, ROCs. The mean ROC quadratic constants for Experiments 1, 2, and 3 are all negative, and each differs from zero at *p* levels less than .001 (see the quadratic constants and their SEs in Table 1). Eleven other data sets show convex ROCs (see Table 2). This type of threshold model does not fit the data.

[10] A confidence rating model based on threshold theory by Erdfelder and Buchner (1998) is concerned with item recognition in a more complex experimental procedure: process dissociation.

[11] Multinomial threshold theory is set up to cover both item and source recognition. Kinchla (1994) pointed out, however, that it has a problem in handling item recognition. He showed that the equations of the full model can be collapsed to focus on item recognition. The resulting equation, however, is the equation for a high threshold process that produces a rectilinear ROC for item recognition. Kinchla pointed out further that there is a wealth of evidence that ROCs for item recognition are curvilinear, not rectilinear (Egan, 1958; Murdock, 1965; Ratcliff et al., 1992). We now follow Kinchla's procedure and collapse the model to examine its implications for source recognition when disentangled from item recognition.

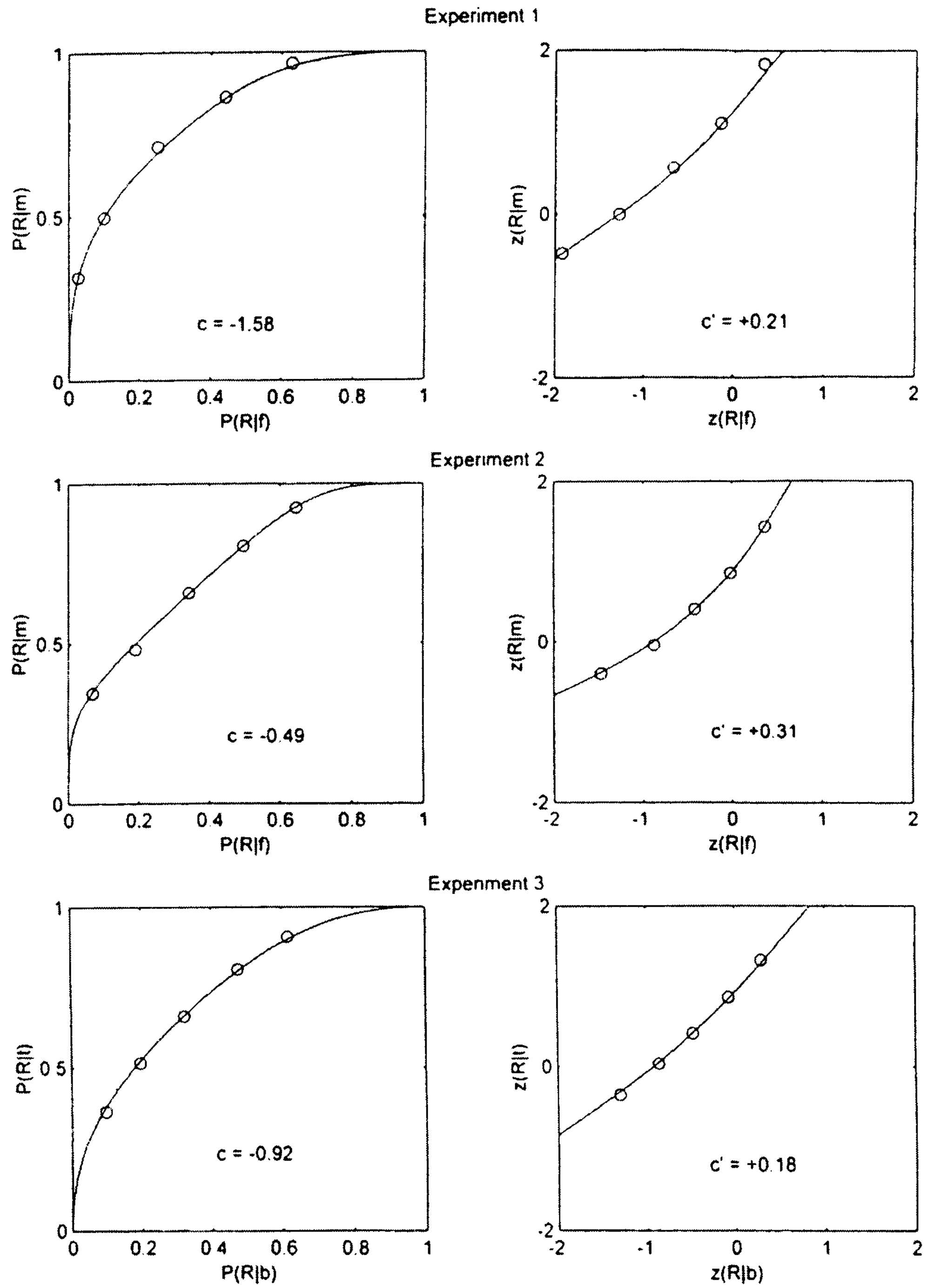

Figure 5. Revised 2D-SDT (two-dimensional signal-detection theory) model fits for the receiver operating characteristic (ROCs) and *z*-ROCs of Experiments 1, 2, and 3.

The preceding discussion has centered on a simple extension of one family of multinomial processing models. In their response to Kinchla (1994), Batchelder, Riefer, and Hu (1994) discussed the possible substitution of low-threshold or signal-detection mechanisms for the high-threshold mechanisms in Batchelder and Riefer (1990). Bayen et al. (1996) applied multinomial processing models based on low-threshold and double high-threshold mechanisms. These also reduce, however, to rectilinear ROCs.

Another approach involving threshold mechanisms has recently been presented by Malmberg (2002). This approach is, however, programmatic, indicating general ways in which other threshold models might be constructed. Fitted models are not constructed or tested. Demonstrations are presented that an expansion of a double high-threshold model can generate the curvilinear ROCs such as those found in item and source recognition. The expansion involves, however, the assumption of a large number of parameters.

For example, the ROCs of Yonelinas's (1999) experiment are displayed and are approximated by using 22 parameters. This is not a fit because the 22 parameters are used to cover 20 observed values. The demonstration shows that fits may be possible but are not actualized at this time. A viable threshold model to which data have been fitted does not exist at this time.

Dual-Process Model

Yonelinas's (1999) dual-process model is a hybrid model that combines threshold and signal-detection components. He has extended the model, originally designed for item recognition memory, to cover source recognition memory. The extension reduces to a simple threshold model for the data presented here. The dual-process model, as presented, assumes that there are two different types of memory information: familiarity and recollection. Familiarity is processed by a signal-detection mechanism. Recollection is processed by a threshold mechanism. This distinction between familiarity and recollection was first proposed by Mandler (1980) and has motivated, in various forms, a large body of research (e.g., Gardiner, 1988; Jacoby, 1991; Jacoby, Toth, & Yonelinas, 1993; McElree, Dolan, & Jacoby, 1999; Rajaram, 1993; Tulving, 1985). Because of the incorporation of that distinction in the dual-process model, however, the dual-process model reduces to a simple threshold model when applied to source memory tasks such as those described in Experiments 1, 2, and 3, in which the source memory tasks are synchronous and matched; that is, in which the two sources do not differ in familiarity.

The dual-process model offers the following two equations:

$$P(R_j|A) = r_A + (1 - r_A)\int_{-\infty}^{c_j} G(x|A)dx, \qquad (14)$$

$$P(R_j|B) = r_B + (1 - r_B)\int_{-\infty}^{c_j} G(x|B)dx, \qquad (15)$$

where r_A is the probability of recollection of an item from Source A; r_B is the probability of recollection of an item from Source B; j is the confidence rating; and G is a normal distribution. The density functions for A and B are assumed to have the same variances and may differ only in their means. Equations 14 and 15 correspond to Yonelinas's Equations 5 and 6. (For ease of comparison, we have changed the letters in Yonelinas's equations: A replaces t, for target; B replaces l, for lure.)

For the analysis of the dual-process model, there is a particularly important case. That is the case in which the two sources are equal in familiarity. We noted earlier that in our experiments the two sources were designed to be equal in familiarity. We noted also that the analyses of the item recognition d's supported the assertion that they were indeed equal in familiarity. When the two sources are equal in familiarity, the dual-process model assumes that the density functions for A and B have the same mean. By assumption also, the two density functions for A and B are both normal and of equal variance. Therefore, $G(x|A) = G(x|B)$.

When $G(x|A) = G(x|B)$, Equations 14 and 15 give the following rectilinear equation:

$$P(R_j|A) = \frac{r_A - r_A r_B - r_B}{1 - r_B} + \frac{(1 - r_A)}{(1 - r_B)} \cdot P(R_j|B). \qquad (16)$$

In other words, when familiarity of sources is the same, the dual-process model reduces to simple high threshold form and the ROC is predicted to be rectilinear with slope $1 - r_A/1 - r_B$ and intercept $r_A - r_A\, r_B - r_B/1 - r_B$.

In all three experiments reported earlier in this article, the sources are of equal familiarity. Therefore, dual-process theory predicts that they should have rectilinear ROCs. All three experiments contradict that prediction and the theory. Slotnick et al. (2000) and Qin et al. (2001) made the same argument on the basis of their experiments.

Yonelinas (1999) argued that source recognition ROCs are rectilinear on the basis of his Experiment 1, summarized earlier. He tested the group ROC of that experiment and, although it had measurable convexity (see Table 2), did not find the curvilinearity statistically significant, $F(1, 2) = 9.08$, $.05 < p < .10$. He concluded that dual-process theory was supported. However, to claim that this borderline finding indicates rectilinearity accepts the null hypothesis. Moreover, his Experiment 2, which did not show any evidence of a "familiarity" difference in the two sources, had an ROC that is also convex (see Footnote 8 and Table 2). A test of its curvilinearity finds $F(1, 2) = 14.64$, $.05 < p < .10$. Combining the results of the two experiments (Edgington, 1972) rejects rectilinearity with a combined p of .02.

It should be noted that in Yonelinas's (1999) Experiments 1 and 2, which do not show differences in familiarity, the *z*-ROC slopes are 1.00 as in our data (see Table 2). In his Experiments 3 and 4, which show such differences, the slopes depart from 1.00 as noted earlier.

In summary, dual-process theory as currently stated needs revision. Sources that do not differ in familiarity generate convex, not rectilinear ROCs. This is borne out unequivocally by our data and the data of Slotnick et al. (2000), Qin et al. (2001), and Yonelinas's (1999) Experiments 1 and 2. Item and source recognition do not fall on a single axis. The revision requires a movement from a one-dimensional to a two-dimensional structure such as that in Figure 2.

Summary

The results of our three experiments show that ROCs for source recognition show three regularities. The ROCs are convex, *z*-ROCs are concave, and linear slopes of *z*-ROCs are near 1.00. These regularities are supported by the findings of other investigators. These regularities contrast with the regularities of item recognition. Item recognition produces ROCs that are convex but *z*-ROCs that are rectilinear with linear slopes near 0.80. The findings on source recognition are based on two different paradigms—one with both item and source recognition tested consecutively (Experiment 1), the other with source recognition tested independently of item recognition (Experiments 2 and 3). The findings are also based on stimuli in two different modalities—auditory (Experiments 1 and 2) and visual (Experiment 3). The consistency of results with these different paradigms and with those of other studies (see Table 2) underlines their generality. Moreover, Experiments 2 and 3 involve a simplified one-dimensional decision space. Experiment 1 involves a two-dimensional space. All give the same pattern of results. The complexities of a two-dimensional decision space that have been considered in discussions of identification (Thomas & Olzak, 1992) do not affect the regularities of concern here.

These results indicate that an initial signal-detection model for source recognition, 2D-SDT, covers a major part of the variance in ROC data but requires supplementing for full fitting to data. An extension using mixture distributions does cover all three regularities fully. The empirical results and conclusions drawn from them are fully concordant with the results of Qin et al. (2001) and Slotnick et al. (2000). They are also concordant with the results of Banks's (2000) analysis of the dimensionality and geometry of the recognition memory space. We have carried the work on that space further here, with formal modeling of the distributions that space contains, on the basis of the regularities of source recognition.

Practical and Theoretical Implications

It was pointed out earlier that there are strong practical implications that come from the model adopted for source recognition (Kinchla, 1994). A difference in performance that in one model would be interpreted as a simple difference in bias or strategy would be interpreted in another as a difference in accuracy or capacity. Moreover, the revised model designates an attention component, λ, which can be measured in a simple recognition performance. That measure permits a fuller analysis of a deficit that occurs with aging or brain damage, that is, whether the deficit reflects impairments of sensitivity or attention or both. There are also further theoretical implications. Both the 2D-SDT model and its revision lead to the prediction that any operation that improves item recognition accuracy (e.g., repetition) for both sources will improve source recognition. This prediction follows from the effects obtained when the triangle of discrimination dimensions seen in Figure 2 is enlarged by moving distributions A and B further out on axis NA and axis NB, corresponding to an increase in item recognition accuracy. This prediction does not derive in any obvious way from available forms of threshold theory or dual-process theory. An opposing view of the relation of source to item memory is, however, presented by Lindsay and Johnson (1991).

In summary, a revised 2D-SDT model is fully adequate for the explanation of the presented source recognition data. As such, the model can be considered for other domains in which source memory is pertinent. Extensions of the model to other experimental paradigms, indicated briefly here, will afford further tests of the model.

References

Ashby, F. G. (1992). *Multidimensional models of perception and cognition.* Hillsdale, NJ: Erlbaum.

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

Batchelder, W. H., Hu, X., & Riefer, D. M. (1994). Analysis of a model for source monitoring. In G. H. Fischer & D. Laming (Eds.), *Contributions to mathematical psychology, psychometrics, and methodology* (pp. 51–65). New York: Springer-Verlag.

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 discrimination, item detection, and multinomial models of source monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 197–215.

Benjamin, A. S., & Craik, F. I. M. (2001). Parallel effects of aging and time pressure on memory for source: Evidence from the spacing effect. *Memory & Cognition, 29,* 691–697.

Bishop, Y. M. M., Fienberg, S. E., & Holland, P. W. (1975). *Discrete multivariate analysis.* Cambridge, MA: MIT Press.

Brown, A. S., Jones, E. M., & Davis, T. L., (1995). Age differences in conversational source monitoring. *Psychology and Aging, 10,* 111–122.

Cohen, G., & Faulkner, D. (1989). Age differences in source forgetting: Effects on reality monitoring and on eyewitness testimony. *Psychology and Aging, 4,* 10–17.

DeCarlo, L. T. (1998). Signal-detection theory and generalized linear models. *Psychological Methods, 3,* 186–205.

DeCarlo, L. T. (2000a, September). *An application of a bivariate probit signal detection model to source monitoring and its implementation as a multiple group structural equation model.* Paper presented at the meeting of the European Mathematical Psychology Group, Graz, Austria.

DeCarlo, L. T. (2000b, August). *An extension of signal detection theory via finite mixtures of generalized linear models.* Paper presented at the annual meeting of Society of Mathematical Psychology, Kingston, Ontario, Canada.

DeCarlo, L. T. (2002). Signal detection theory with finite mixture distributions: Theoretical developments with applications to recognition memory. *Psychological Review, 109,* 710–721.

Degli'Innocenti, A., & Bäckman, L. (1996). Aging and source memory: Influences of intention to remember and associations with frontal lobe tests. *Aging, Neuropsychology, and Cognition, 3,* 307–319.

Dywan, J., & Jacoby, L. L., (1990). Effects of aging in source monitoring: Differences in susceptibility to false fame. *Psychology and Aging, 5,* 379–387.

Dywan, J., Segalowitz, S. J., & Webster, L. (1998). Source monitoring: ERP evidence for greater reactivity to nontarget information in older adults. *Brain and Cognition, 36,* 390–430.

Edgington, E. S. (1972). An additive method for combining probability values from independent experiments. *Journal of Psychology, 80,* 351–365.

Egan, J. P. (1958). *Recognition memory and the operating characteristic* (Tech. Rep. No. AFCRC-TN-58-51). Bloomington: Indiana University Hearing and Communication Laboratory.

Egan, J. P. (1975). *Signal detection theory and ROC analysis.* New York: Academic Press.

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

Ferguson, S. A., Hashtroudi, S., & Johnson, M. K. (1992). Age differences in using source-relevant cues. *Psychology and Aging, 7,* 443–452.

Foley, M. A., Johnson, M. K., & Raye, C. L. (1983). Age-related changes in confusion between memories for thoughts and memories for speech. *Child Development, 54,* 51–60.

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

Geiselman, R. E., & Bellezza, F. S. (1976). Long-term memory for speaker's voice and source location. *Memory & Cognition, 4,* 483–489.

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

Glanzer, M., Adams, J. K., Iverson, G. J., & Kim, K. (1993). The regularities of recognition memory. *Psychological Review, 100,* 546–567.

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

Glanzer, M., Kim, K., Hilford, A., & Adams, J. K. (1999). Slope of the receiver-operating characteristic in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 500–513.

Glisky, E. L., Polster, M. R., & Routhieaux, B. C. (1995). Double dissociation between item and source memory. *Neuropsychology, 9,* 229–235.

Green, D. M., & Swets, J. A. (1974). *Signal detection and psychophysics.* New York: Krieger.

Harvey, P. D. (1985). Reality monitoring in mania and schizophrenia. *Journal of Nervous and Mental Disease, 173,* 67–72.

Hashtroudi, S., Johnson, M. K., & Chrosniak, L. D. (1989). Aging and source monitoring. *Psychology and Aging, 4,* 106–112.

Henkel, L. A., Johnson, M. K., & DeLeonardis, D. M. (1998). Aging and source monitoring: Cognitive processes and neuropsychological correlates. *Journal of Experimental Psychology: General, 127,* 251–268.

Hintzman, D. L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review, 95,* 528–551.

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

Jacoby, L. L., Toth, J. P., & Yonelinas, A. P. (1993). Separating conscious and unconscious influences of memory: Measuring recollection. *Journal of Experimental Psychology: General, 122,* 139–154.

Janowsky, J. S., Shimamura, A. P., & Squire, L. R. (1989). Source memory impairment in patients with frontal lobe lesions. *Neuropsychologia, 27,* 1043–1056.

Johnson, M. K., DeLeonardis, D. M., Hashtroudi, S., & Ferguson, S. A. (1995). Aging and single versus multiple cues in source monitoring. *Psychology and Aging, 10,* 507–517.

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

Johnson, M. K., Kounios, J., & Nolde, S. F. (1996). Electrophysiological brain activity and memory source monitoring. *NeuroReport, 7,* 2929–2932.

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.

Johnson, M. K., & Raye, C. L. (1981). Reality monitoring. *Psychological Review, 88,* 67–85.

Kausler, D. H., & Puckett, J. M. (1980). Adult age differences in recognition memory for a nonsemantic attribute. *Experimental Aging Research, 6,* 349–355.

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

Krantz, D. H. (1969). Threshold theories of signal detection. *Psychological Review, 76,* 308–324.

Lindsay, D. S., & Johnson, M. K. (1989). The eyewitness suggestibility effect and memory for source. *Memory & Cognition, 17,* 349–358.

Lindsay, D. S., & Johnson, M. K. (1991). Recognition memory and source monitoring. *Bulletin of the Psychonomic Society, 29,* 203–205.

Lindsay, D. S., Johnson, M. K., & Kwon, P. (1991). Developmental changes in memory source monitoring. *Journal of Experimental Child Psychology, 52,* 297–318.

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

Malmberg, K. J. (2002). On the form of ROCs constructed from confidence ratings. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 380–387.

Mandler, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological Review, 87,* 252–271.

Mather, M., Johnson, M. K., & DeLeonardis, D. M. (1999). Stereotype reliance in source monitoring: Age differences and neuropsychological test correlates. *Cognitive Neuropsychology, 16,* 437–458.

Matlab. (1995). Natick, MA: Mathworks, Inc.

McElree, B., Dolan, P. O., & Jacoby, L. J. (1999). Isolating the contributions of familiarity and source information to item recognition: A time course analysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 563–382.

McIntyre, J. S., & Craik, F. I. M. (1987). Age differences in memory for item and source information. *Canadian Journal of Psychology, 41,* 175–192.

Mitchell, D. B., Hunt, R. R., & Schmitt, F. A. (1986). The generation effect and reality monitoring: Evidence from dementia and normal aging. *Journal of Gerontology, 41,* 79–84.

Murdock, B. B., Jr. (1965). Signal-detection theory and short-term memory. *Journal of Experimental Psychology, 70,* 443–447.

Murdock, B. B., Jr. (1974). *Human memory: Theory and data.* Potomac, MD: Erlbaum.

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

Paivio, A., Yuille, J. C., & Madigan, S. A. (1968). Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology: Monograph Supplement, 76*(1, Pt. 2), 1–25.

Qin, J., Raye, C. L., Johnson, M. K., & Mitchell, K. J. (2001). Source ROCs are (typically) curvilinear: Comment on Yonelinas (1999). *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* 1110–1115.

Rajaram, S. (1993). Remembering and knowing: Two means of access to the personal past. *Memory & Cognition, 21,* 89–102.

Ratcliff, R., McKoon, G., & Tindall, M. (1994). Empirical generality of data from recognition memory receiver-operating characteristic functions and implications for the global memory models. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 763–785.

Ratcliff, R., Sheu, C.-F., & Gronlund, S. D. (1992). Testing global memory models using ROC curves. *Psychological Review, 99,* 518–535.

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.

Rose, R. G., Rose, P. R., King, N., & Perez, A. (1975). Bilingual memory for related and unrelated sentences. *Journal of Experimental Psychology: Human Learning and Memory, 1,* 599–606.

Rotello, C. M., Macmillan, N. A., & Van Tassell, G. (2000). Recall-to-reject in recognition: Evidence from ROC curves. *Journal of Memory and Language, 43,* 67–85.

Saegert, J., Hamayan, E., & Ahmar, H. (1975). Memory for language of input in polyglots. *Journal of Experimental Psychology: Human Learning and Memory, 1,* 607–613.

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.

Schacter, D. L, Kaszniak, A. W., Kihlstrom, J. F., & Valdiserri, M. (1991). The relation between source memory and aging. *Psychology and Aging, 6,* 559–568.

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.

Shimamura, A. P., & Squire, L. R. (1991). The relationship between fact and source memory: Findings from amnesic patients and normal subjects. *Psychobiology, 19,* 1–10.

Slotnick, S. D., Klein, S. A., Dodson, C. S., & Shimamura, A. P. (2000). An analysis of signal detection and threshold models of source memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 1499–1517.

Swets, J. A., Tanner, W. P., Jr., & Birdsall, T. G. (1961). Decision processes in perception. *Psychological Review, 68,* 301–340.

Tanner, W. P., Jr. (1956). Theory of recognition. *Journal of the Acoustical Society of America, 28,* 882–888.

Thomas, J. P., & Olzak, L. A., (1992). Simultaneous detection and identification. In F. G. Ashby (Ed.), *Multidimensional models of perception and cognition* (pp. 253–277). Hillsdale, NJ: Erlbaum

Tulving, E. (1985). Memory and consciousness. *Canadian Psychology, 26,* 1–12.

Vermunt, J. K. (1997). *LEM: A general program for analysis of categorical data.* Tilburg, The Netherlands: Tilburg University.

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

Yonelinas, A. P. (1997). Recognition memory ROCs for item and associative information: The contribution of recollection and familiarity. *Memory & Cognition, 25,* 747–763.

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

Yonelinas, A. P. (2001). Consciousness, control, and confidence: The 3 Cs of recognition memory. *Journal of Experimental Psychology: General, 130,* 361–379.

Yonelinas, A. P., Dobbins, I., Szymanski, M. D., Dhaliwal, H. S., & King, L. (1996). Signal-detection, threshold, and dual-process models of recognition memory: ROCs and conscious recollection. *Consciousness and Cognition, 5,* 418–441.

Yonelinas, A. P., Kroll, N. E. A., Dobbins, I., Lazzara, M., & Knight, R. T. (1998). Recollection and familiarity deficits in amnesia: Convergence of remember-know, process dissociation, and receiver operating characteristic data. *Neuropsychology, 12,* 323–339.

Zaragoza, M. S., & Lane, S. (1991, November). *The role of attentional resources in suggestibility and source monitoring.* Paper presented at the 32nd annual meeting of the Psychonomics Society, San Francisco.

Appendix

Number of Responses in Each Confidence Rating for Each of the Experiments Summed Across Participants

Experiment and participant	Rating 1	2	3	4	5	6
Experiment 1						
Male	264	133	159	116	80	33
Female	25	62	105	128	145	294
Experiment 2						
Male	1,358	550	701	585	466	300
Female	278	477	598	614	589	1,404
Experiment 3						
Top	1,313	538	524	529	363	333
Bottom	348	356	455	552	510	1,379

Received January 22, 2002
Revision received August 6, 2002
Accepted August 6, 2002 ■