Journal of Experimental Psychology:
Learning, Memory, and Cognition
2004, Vol. 30, No. 6, 1176–1195

0278-7393/04/$12.00 DOI: 10.1037/0278-7393.30.6.1176

Six Regularities of Source Recognition

Murray Glanzer and Andy Hilford
New York University

Kisok Kim
Iona College

In recent work, researchers have shown that source-recognition memory can be incorporated in an extended signal detection model that covers both it and item-recognition memory (A. Hilford, M. Glanzer, K. Kim, & L. T. DeCarlo, 2002). In 5 experiments, using learning variables that have an established effect on item recognition, the authors tested further implications of that extended model. The results establish 6 source-recognition regularities that support the model. An important regularity in that set is the linkage of source and item recognition: Any learning variable that increases item recognition increases source recognition.

Item recognition is remembering an item's occurrence (e.g., whether or not a story was heard at a party, whether or not a word was seen on a study list). Source recognition is remembering the source of the item (e.g., whether the story was heard at a party or on the radio, whether a word was spoken by a man or a woman). Source recognition has become a major area of study as a result of the pioneering work of Johnson (Johnson, Hashtroudi, & Lindsay, 1993; Johnson & Raye, 1981) on reality monitoring, Schacter (Schacter, Kaszniak, Kihlstrom, & Valdiserri, 1991) on aging and Shimamura and Squire (1987, 1991) on amnesia.

Item recognition has been fully incorporated in signal detection theory (SDT) models (Egan, 1958; Gillund & Shiffrin, 1984; Glanzer, Adams, Iverson, & Kim, 1993; Green & Swets, 1974; Murdock, 1965; Murdock, 1974; Ratcliff, McKoon, & Tindall, 1994; Ratcliff, Sheu, & Gronlund, 1992; Shiffrin & Steyvers, 1997) either by SDT in isolation (e.g., Egan, 1958; Murdock, 1965) or by SDT incorporated in other models (Gillund & Shiffrin, 1984; Glanzer et al., 1993; Murdock, 1982; Shiffrin & Steyvers, 1997). The use of SDT is validated by five regularities displayed by item-recognition data. The first four regularities, listed next, are those seen in receiver operating characteristics (ROCs).

1. The ROCs are convex (Egan, 1958; Lawrence & Banks, 1973; Murdock, 1965, 1974).

2. The z-ROCs (ROCs plotted in z-scores, standard scores) are linear (Allen & Garton, 1968; Glanzer, Hilford, Kim, & Adams, 1999; Glanzer, Kim, Hilford, & Adams, 1999; Ratcliff et al., 1994).

3. The slopes of the z-ROCs are less than 1.00 (Glanzer, Kim, et al., 1999; Ratcliff et al., 1994).

4. The slopes of the z-ROCs change with accuracy, declining as accuracy increases. (Glanzer & Adams, 1990; Glanzer et al., 1993; Glanzer, Kim, et al., 1999).

5. The underlying distributions show mirror effects (Glanzer & Adams, 1985, 1990; Hilford, Glanzer, & Kim, 1997; Kim & Glanzer, 1993). When there are two conditions that produce differences in accuracy, the underlying old (O) distributions and new (N) distributions array themselves in the following pattern:

$$SN < WN < WO < SO,$$

where S = strong, W = weak, N = new, and O = old. This order, which defines the mirror effect, can also be seen in the array of hits and false alarms in yes–no data or in the means of confidence rating data.

The first regularity, convex ROCs, implies continuous or multiple-valued underlying distributions and rules out some threshold models. The second, linear z-ROCs, supports an SDT model in which underlying normal or approximately normal distributions are assumed. The third, z-ROC slopes less than 1.00, supports an SDT model in which the distributions of old (studied) items have larger variance than the distributions of new items. The justification for the variance difference may be phrased as follows. In item recognition, the old items have been exposed to an additional, varied effect, in other words, the effect of the study session. This additional effect increases the variance of the old distribution. When a larger variance distribution is plotted against a smaller variance distribution in a z-ROC, the slope declines from 1.00. The fourth regularity, decreasing slopes with increasing accuracy, follows from the preceding statement. For example, when the variable is repetition, the first study trial produces the larger variance of the old distribution, as compared with the new distribution and, therefore, a z-ROC slope less than 1.00. For some contrary results, however, see Heathcote (2003). A second study trial will increase

---

Murray Glanzer and Andy Hilford, Department of Psychology, New York University; Kisok Kim, Department of Psychology, Iona College.

The research reported in this article was carried out under National Institute of Mental Health Grant 1R01 MH60114. We thank Larry Maloney for his valuable advice. We also thank Simonne Pollini for her excellent work testing participants.

Correspondence concerning this article should be addressed to Murray Glanzer or Andy Hilford, Department of Psychology, New York University, 6 Washington Place, New York, NY 10003. E-mail: mg3@nyu.edu or ah13@nyu.edu

variance of the old distribution still further and lower the slope of the z-ROC further.

The fifth regularity, mirror effects, is explained within the context of the theory by invoking likelihood ratios, which are the basis of SDT decision axes (although not generally used in psychological work). See Green and Swets (1974) and Thomas (1985). These likelihood ratios have been brought to the fore in recent work on memory (Anderson & Milson, 1989; Dennis & Humphreys, 2001; Glanzer et al., 1993; McClelland & Chappell, 1998; Shiffrin & Steyvers, 1997).

For a detailed example of the joining of SDT with a memory theory that results in predictions of these regularities, see Glanzer et al. (1993). That article details the process and representation that is assumed here for both item and source recognition. There have, of course, been other theories that join with SDT to address characteristics of item recognition, for example, Shiffrin and Steyvers (1997).

Recent empirical and theoretical work has been directed at an expanded theory of memory processes (Ashby, 1992) and, specifically, at incorporating source recognition in multidimensional SDT (Banks, 2000; Hilford et al., 2002; Slotnick, Klein, Dodson, & Shimamura, 2000). In this article, we aim to advance that incorporation by establishing the regularities of source recognition that support a SDT model. The regularities parallel those of item recognition.

Hilford et al. (2002) examined a two-dimensional signal detection model (2D-SDT) for source and item recognition using ROCs. The model was adapted from a model originally presented by Tanner (1956) for sensory discrimination. For a further analysis of the Tanner model, see Duncan (1999). The 2D-SDT model, a simple extension of the standard model for item recognition, is depicted in Figure 1. We have changed the labels of the distributions from those used by Tanner to translate the model to source and item recognition with the following labels: N for new items, A for old items from source A, and B for old items from source B. The ordinate of this plot is the item-recognition dimension. The abscissa is the source-recognition dimension. The distances between means represent the accuracy of discrimination among the three distributions. The larger the distance, d, the more accurate the discrimination. Item-recognition accuracy is represented by either d(AN) or d(BN). Source-recognition accuracy is represented by d(AB).

In the upper row of Figure 2, we reduce the general two-dimensional plot in Figure 1 to plots of means and distances between means. The first plot, I, represents a case in which the two item-recognition distances d(AN) and d(BN) and the source-recognition distance d(AB) are all equal. The means form an equilateral triangle. The second case, II, forms an isosceles triangle. It represents the case in which the two item-recognition distances d(AN) and d(BN) are equal but the source-recognition d(AB) is less than the item-recognition distance; in other words, source-recognition accuracy is less than item-recognition accuracy. Examples of such a case are found in the experiments of Slotnick et al. (2000). The third case, III, represents the case in which item recognition for the two sources is not equal, d(AN) $\neq$ d(BN). The means form a scalene triangle. An example of such a case is found in Yonelinas (1999). We also discuss another example of Case III in detail later in this article (Lindsay & Johnson, 1991).

The theoretical assertions with respect to study and test implicit in this model are the following. Initial presentation of an item produces information that has relevance to both item and source recognition. This is the same assumption as that made by Johnson et al. (1993; see their Figure 1). This point and its SDT implications for a structure like that in Figure 1 have been spelled out by Thomas (1985) for sensory detection and identification, the homologues of item recognition and source recognition in memory:

> . . . the two abilities are closely interrelated. The relationship arises from a common dependence on the manner in which the . . . system encodes or forms neural representation.. One way to describe the relationship between detection and identification is that they represent different uses of the same . . . information. They involve the application of different decision processes to the same neural representation . . . The performance of both tasks depends on the strength or intensity of the stimuli . . . [References specific to the visual system have been eliminated from the quotation to generalize the SDT structure to cover memory.] (p. 1457)

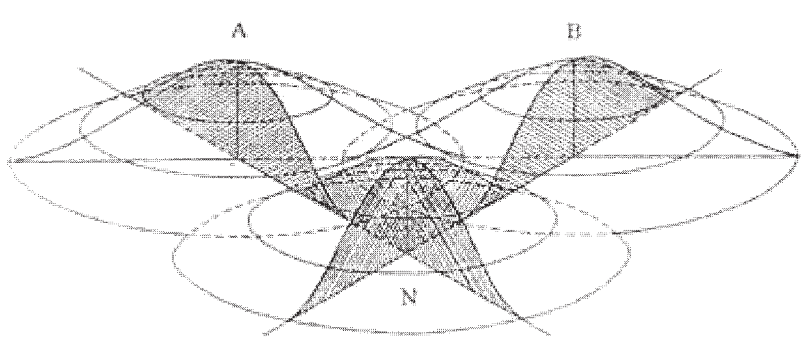

Figure 1. The two-dimensional signal detection model for both item recognition and source recognition adapted from Tanner's (1956) model for sensory discrimination. The means for the new distribution, N, and the two old distributions from source A and source B are indicated on the decision axes. From "Theory of Recognition," by W. P. Tanner, Jr., 1956, *Journal of the Acoustical Society of America*, *28*, p. 883. Copyright 1956 by the Acoustical Society of America. Adapted with permission.

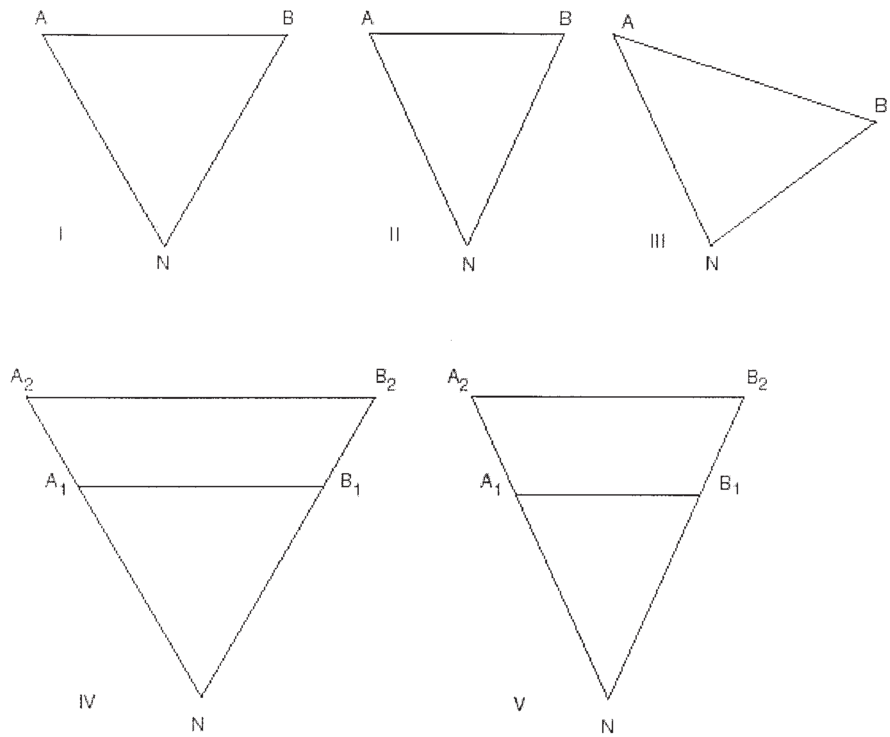

Figure 2. Simplified two-dimensional signal detection theory structures for item and source recognition. Only the means for the two sources of old items, A and B, and the new items, N, and the distances between them are represented. Panel I: Item recognition for A and B and source recognition all equal. Panel II: Item recognition for A and B equal and better than source recognition. Panel III: Item recognition for A and B unequal. Panel IV: The effect of increasing item recognition on Panel I. A_1 and B_1 represent the effect of the initial item-recognition levels, and A_2 and B_2 represent the effect of the increased levels. Panel IV: The effect of increasing item recognition on Panel II.

This generalization has three implications:

1. On test, item-recognition decisions are governed by SDT mechanisms. Therefore the item-recognition tests should produce regular ROCs with the regularities noted earlier.

2. On test, source-recognition decisions are also governed by SDT mechanisms. Therefore, source-recognition tests should also produce regular ROCs with a corresponding set of regularities.

3. In the case of memory, "strength", noted in the quotation, is changed by factors that affect learning. Therefore, any effective learning variable should produce linked, correlated effects on item and source recognition. In other words, any variable that improves item recognition should improve source recognition. The testing of this implication is one of our purposes in this study.

The preceding statement stems from a model in which the structure depicted in Figure 1 is assumed and, further, in which it is assumed that item learning moves the means of distributions A and B further out on the decision axes AN and BN. The possible arrangement of means with two levels of a variable that affects item recognition are shown in the lower row of Figure 2. A_1 and B_1 represent means for the less effective item-recognition condition. A_2 and B_2 represent the means for the more effective condition. When learning produces equivalent effects on item and source recognition, two overlapping equilateral triangles are obtained, as shown in IV. When the learning involving source information is less effective than the learning involving item information, then the means form two overlapping isosceles triangles, as shown in V. In both cases, source recognition in the better item-recognition condition is greater than the source-recognition accuracy in the less accurate condition: $d(A_2B_2) > d(A_1B_1)$. The model predicts that item-recognition accuracy and source-recognition accuracy are linked. This is a prediction that follows from the system outlined here.

There is, as we noted earlier, considerable evidence supporting the first implication: that item recognition is effectively covered by SDT. There is also evidence supporting the second: that source recognition is effectively covered by SDT. In three experiments, Hilford et al. (2002) obtained source-recognition ROCs that showed three regularities:

1. Source-recognition ROCs were convex. This supports an

SDT model with the usual assumption of underlying continuous distributions.

2. Slopes of z-ROCs (the ROCs plotted on z-score axes) were close to 1.00. This supports an SDT model with underlying distributions of equal variance. This differs from the parallel regularity in item recognition. One may explain it by noting that in source recognition, the underlying distributions both consist of old items. The distributions do not have different histories and, therefore, do not differ in variance.

3. The z-ROCs were slightly concave. This departure from linearity indicated that some addition to the model was needed. We covered the small deviation by supplementing the model, assuming that some items' source information is not attended to. We represented this in a modified SDT model by assuming that the decision axis for source recognition has, in addition to the two distributions A and B on the AB axis in Figure 1, a third distribution N′ whose mean is zero. This additional distribution produces two mixture distributions A + N′ and B + N′, which are the basis for source decisions. For further details on mixture models, see DeCarlo (2002) and Hilford et al. (2002). In the present article, we do not consider this theoretical extension to cover the concavity further, because the departure from rectilinearity is slight.

The source ROC data are, however, well fitted by the simpler SDT model (without the extension to a mixture model). Thus in the three experiments of Hilford et al. (2002), the fit of the simpler model accounts for 98%, 99%, and 99% of the data. We therefore use the simpler model here, focusing on the major effects. We fit the model as in Hilford et al. by linear regression. The R^2s give a measure of adequacy of fit that indicates that the model accounts again for an average of 99% of the data. We do, however, include tests of the concavity to further demonstrate the reliability of that small effect.

The three regularities—convex ROCs, z-ROC slopes of 1.00, and concave z-ROCs—have also been found in source-recognition ROCs of Qin, Raye, Johnson, and Mitchell (2001), Slotnick et al. (2000), and Yonelinas (1999). For a survey of those concordant findings, see Hilford et al. (2002).

The experiments that follow replicated the findings on the first three source-recognition regularities. In addition, because we varied item-recognition accuracy, they also established a source-recognition regularity, slope constancy, that contrasts with the fourth item-recognition regularity, slope change. The experiments also demonstrated a source-recognition regularity that is the same as the fifth item-recognition regularity, the mirror effect. Finally, and most important, they permitted us to establish a sixth regularity, the linkage of source-recognition accuracy to item-recognition accuracy, that changes in item-recognition accuracy produce corresponding changes in source-recognition accuracy. We applied five variables that are known to affect item recognition. Because the 2D-SDT model implies that learning affects both the item information and source information, variables that increase item-recognition accuracy should also increase source-recognition accuracy.

We noted earlier the item regularity of slope changes with changes in accuracy. We already know that source recognition has a contrasting regularity of slope stability despite changes in accuracy across experiments. We know this because there are several experiments (see Hilford et al., 2002) that show z-ROCs with slopes of 1.00 despite variation in accuracy between experiments. We now show that stability within experiments. For source recognition, slopes should stay constant at 1.00. The reason for this stability is that, as noted earlier, the slopes reflect the histories of the underlying distributions. In item recognition, the two underlying distributions, old and new, have different histories. In source recognition, the two relevant underlying distributions are both old, do not differ in history, and, therefore, do not differ in variance. This is true, of course, only when, as in the present experiments and others (Hilford et al., 2002; Slotnick et al., 2000), the two sources are given equivalent treatment, for example, the same study time per item. If a variable is then introduced that is applied to items from both sources equally, the two underlying distributions will again have matching histories and matching variances. In other words, two sources' distributions will have equal variance when the accuracy of item recognition is equal and low for both (e.g., A_1 and B_1 in IV and also V of Figure 2), and when accuracy of item recognition is equal and high for both sources' distribution (e.g., A_2 and B_2 in IV and V of Figure 2). Therefore, no changes in slope are expected.

The prediction of an increase in source-recognition accuracy with an increase in item-recognition accuracy stems from the assumption, implicit in Figure 2, that the distances exist in a Euclidean space, and the assumption that any change in accuracy means a movement of distributions on the discrimination axis. In earlier work (Glanzer & Adams, 1990; Glanzer et al., 1993), we demonstrated a mirror effect in item recognition. When there are two experimental conditions, a weak condition, W, that produces relatively low accuracy and a strong condition, S, that produces relatively high accuracy, then the proportion P of hits ("yes" to old items, O) and false alarms ("yes" to new items, N) for the two conditions will produce a symmetric pattern.

$$\mathrm{P(SN)} < \mathrm{P(WN)} < \mathrm{P(WO)} < \mathrm{P(SO)}$$

In the case of confidence ratings, R, the equivalent mirror pattern on mean ratings is

$$\mathrm{R(SN)} < \mathrm{R(WN)} < \mathrm{R(WO)} < \mathrm{R(SO)}$$

where ratings go from 1 *for sure item is new* to, for example, 6 for *sure item is old*. We explained this pattern by assuming that individuals made recognition decisions on the basis of a likelihood evaluation, deciding whether the given test item is more or less likely to be old or new. We detailed the formal theory and fitted it to data (Glanzer et al., 1993). We extend the assumption of likelihood decision to the source-decision axes of Figures 1 and 2, and therefore predict that when there are two conditions, one weak and one strong, and two sources, A and B, then the confidence ratings will also display a mirror pattern:

$$\mathrm{R}(\mathrm{A}_2) < \mathrm{R}(\mathrm{A}_1) < \mathrm{R}(\mathrm{B}_1) < \mathrm{R}(\mathrm{B}_2)$$

where the ratings go from 1 for *sure item is from source* A to, for example, 6 for *sure item is from source* B.

The sixth possible regularity, the relation expected between item-recognition accuracy and source-recognition accuracy, is also implicit in the 2D-SDT model presented in Figure 1 and in the analyses of the relation between item and source recognition as noted by Thomas (1985). In the 2D-SDT model, any variable that affects acquisition or retention of item-recognition information will affect the acquisition or retention of source-recognition information in the same way. Therefore, any variable that increases item-recognition accuracy should also increase source-recognition accuracy. That is, item-recognition and source-recognition accuracy should be linked. This relation is depicted in cases IV and V of Figure 2.

The literature has reports of three variables that have linked effects on source-recognition and item-recognition accuracy: aging, delay, and divided attention.

1. Aging, which lowers item-recognition accuracy, also lowers source-recognition accuracy (Hashtroudi, Johnson, & Chrosniak, 1989; Spencer & Raz, 1995).

2. Delay, which lowers item-recognition accuracy, also lowers source-recognition accuracy. Bornstein and LeCompte (1995) showed that item recognition and source recognition declined over delay in lockstep. In their Experiment 1, the correlation of four pairs of item- and source-accuracy scores, generated by four different delays, was 1.00.

3. Divided attention, which lowers item-recognition accuracy, also lowers source-recognition accuracy. Troyer, Winocur, Craik, and Moscovitch (1999) carried out two experiments, each with three attention conditions during study—one with full attention and two with divided attention tasks that varied in degree of distraction. The correlation of item- and source-recognition accuracy, based on three pairs of means in each experiment, is perfect in both experiments.

We extended the examination of linkage here with five variables that have established effects on item-recognition accuracy, observing the effects of these variables on source recognition. If there is linkage, these variables should produce homologous effects on source recognition. We, moreover, took the examination of linkage a step further. We also examined the linkage at an individual level, the correlation of individual item, and source-recognition accuracy within each experimental condition. The evidence supporting Regularities 1, 2, and 4 replicated preceding findings (Hilford et al., 2002). Evidence on Regularities 3, 5, and 6 tested the extension of the 2D-SDT model to the structures shown in IV or V in Figure 2.

To summarize, we examine the data of the following experiments for evidence of six regularities.

1. Convex ROCs

2. Slightly concave z-ROCs

3. Constant z-ROC linear slopes

4. z-ROCs with linear slopes of 1.00

5. Mirror effects

6. Linkage of accuracy changes in item and source recognition

The existence of such regularities is important, because they form the basis for a complete theory of recognition. We examine the regularities in experiments involving changes in accuracy. Regularities 3, 5, and 6, to be tested, require changes in accuracy. Those changes are induced in the experiments reported here.

General Procedure

In the five experiments that follow, we used variables that have a well established effect on item-recognition accuracy: repetition, list length, encoding, word frequency, and aging. We constructed all study and test lists by random selection, individually for each participant. In statistical tests, a *p* level of .05 or less defines statistical significance.

In the presentation of the results for each experiment, we carry out the following steps. We first present an overall picture of the ROCs and z-ROCs on the basis of the group data. We used the sum of the responses of all participants to each confidence rating to generate a group ROC and z-ROC. These group ROCs and z-ROCs display much of the information about the regularities in visual form, for example, the convexity of the ROC. Regression analysis of the group ROCs and z-ROCs furnishes information in the form of numerical indices, for example, quadratic constants. Linear regression of the group z-ROCs furnishes the key fitting of the 2D-SDT model to the data. The adequacy of the fits is indicated by the R^2s, the variance accounted for. For complete statistical analysis, we computed the ROCs and z-ROCs for each participant. We then analyzed them to yield various statistics, for example, slopes. We then subjected the means of those statistics to statistical analysis.

Using fitted quadratic equations, we obtained indices of convexity, the negative quadratic constants, from each individual's ROCs. We used the means of those constants to further test Regularity 1 (convex ROCs). Using fitted linear equations, we obtained slopes and intercepts (d'_2, an index of accuracy) from each individual's z-ROCs. We used the mean slopes to evaluate Regularities 3 and 4 (constant z-ROC slopes of 1.00). We used the mean d'_2s to evaluate Regularity 6 (linkage of item and source accuracy). We averaged the mean confidence ratings to test for the presence of mirror effect, Regularity 5. We also obtained measures of concavity to replicate Regularity 2 from each individual's z-ROCs, using fitted quadratic equations.

We used three different procedures in the five experiments. In Experiments 1 and 4, the test lists included both new and old items. The participants responded first whether each item was new or old and then, for items they designated as old, they indicated from which source it came. In Experiment 2 and 3, a pure source test was given after the study list. The participants were given a list consisting of only old items, so defined for the participants, and they were required to indicate from which source each item came. This source-recognition test was followed by a short yes–no item-recognition test. In Experiments 1, 3, and 5, the two sources were male and female voices. In Experiments 2 and 4, the two sources were positions on the computer monitor. These variations on procedure were used in Hilford et al. (2002) and gave concordant results, as they do here. They all produce the same regularities of the source-recognition ROCs and z-ROCs, supporting the generality of the regularities. Experiment 5 differed from the other four experiments in that it included a full, confidence rating test of item recognition as well as source recognition. We used two different study lists, one to measure item recognition and the other to measure source recognition.

The sources (male and female voices, or top and bottom of the screen) were constructed in these experiments to be matched. They did not differ in modality (e.g., heard vs. read or seen vs. imagined). They did not differ in study time per item. The success of the matching is indicated by the fact that item-recognition accuracy for the two sources in each experiment, with one exception noted in Experiment 5, did not differ. This equivalence is checked for each experiment. Differences in the item-recognition accuracy of the sources introduce complexity in the analysis of the effects that were being studied. We would be dealing with Panel III in Figure 2, not I, II, IV, or V. We do, however, discuss later in this article an experiment in which those differences exist. That discussion is postponed until after we demonstrate the regularities for simpler paradigms.

Experiment 1: Repetition

A variable that reliably affects item recognition is repetition (Egan, 1958; Glanzer, Kim, et al., 1999; Hilford et al., 1997; Ratcliff & Murdock, 1976; Ratcliff et al., 1992). Items that are repeated in a study list are more accurately recognized than items that are not. 2D-SDT predicts, therefore, that repetition should also increase source-recognition accuracy.

Method

The participants heard a study list of items. They then saw a word list that tested both item and source recognition. The study list consisted of both single-presentation words and words that were repeated. Half of the single and half of the repeated words were spoken by a man; the others were spoken by a woman. Selection of study-list words, selection of repeated words, selection of words' voicing, and order of both study and test words were randomized individually for each participant. (This meant that the repeated words were randomly separated from each other.) Following the study lists, participants were given a yes–no recognition test on a word list consisting of the studied words and an equal number of new words, distractors. If they responded to a word as "old," they then indicated the source (*man* or *woman*) of the word.

Participants. Seventy New York University undergraduates participated. The student participants in this and the following experiments had been speaking English since the age of 10 or earlier. They all participated to fulfill a psychology class requirement. This description holds for all student participants in Experiments 1–5.

Materials. We chose 360 words from Paivio, Yuille, and Madigan (1968). Two sets of digital recordings of these words were made, one using a male speaker and the other a female speaker, and stored as individual sound files. Forty-five additional words were recorded in the same way for practice and buffer items. The study and test lists were drawn from the main list of 720 words. Both single and repeated items were randomly selected for each list and contained equal numbers of words spoken by the man and the woman.

Procedure. Participants heard study lists (practice and main) presented through headphones connected to a computer. While the study words were presented, the computer screen was blank. The experiment began with a 5-word auditory practice study list followed by a 10-word visual practice test list. After the practice study and test, participants heard the main list—90 single-presentation words and 90 repeated words, a total of 270 words (plus 10 filler words). Half of the single and half of the repeated words were read in a man's voice. The other halves were read in a woman's voice. Repeated words were in the same male or female voice. A 1,000-ms interval separated the words. Participants were told to attend to the words carefully and that they would be tested on them after the completion of the study list. The main study list began and ended with 5 unscored filler items, and the test list was preceded by 5 unscored filler items to eliminate primacy and recency effects. Instructions appeared on the screen before each study or test list to remind participants of their tasks for each upcoming list.

After they heard the study list, the participants were given a combined item- and source-recognition test. Each test list consisted of the original study words plus an equal number of distractors, randomly mixed. These words were presented visually, centered in the middle of the monitor in uppercase letters. For each word, the participant indicated whether the word was old or new by pressing a key marked O for old or a key marked N for new. If the participant responded O, a second display appeared asking the participant to indicate whether the item had been read by the man or the woman. Participants responded by using one of six keys indicating the confidence of their decisions. The confidence ratings were as follows: 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*. The numbers and their descriptions appeared on the monitor. The test was self-paced, with the next test item appearing as soon as the participant responded.

Results

We present recognition data for Experiments 1–5 in two steps. First, to give an overall picture, we show the group ROC, summing across all 70 participants, and the corresponding group z-ROC in plots. The data on which these ROCs and z-ROCs are based are presented in Appendix A. Statistics based on these ROCs and z-ROCs are presented in Appendix B. The group ROC and z-ROC statistics, in all five experiments, conform well to the mean statistics based on the individual data.

After presenting the group data, we summarize and analyze the statistics based on the individual participants' ROCs and z-ROCs. These statistics permit a more complete, detailed analysis.

Group data. The group ROCs and z-ROCs for Experiment 1 are shown in the top panels of Figure 3. The ROCs are convex. The ROCs for the two presentation conditions are separated, with the repeated-presentation ROC higher than single-presentation ROC, indicating greater accuracy with repetition. The group z-ROCs for the two conditions are also separated, indicating, again, a difference in accuracy. The linear slopes of the z-ROCs are near 1.00. All of these statements are supported by the summary statistics given in Appendix B, for example, negative quadratics of the ROC indicating convexity. All of the findings replicate previous findings for source recognition (Hilford et al., 2002).

Another statistic of interest is the R^2, the index of linear fit to the

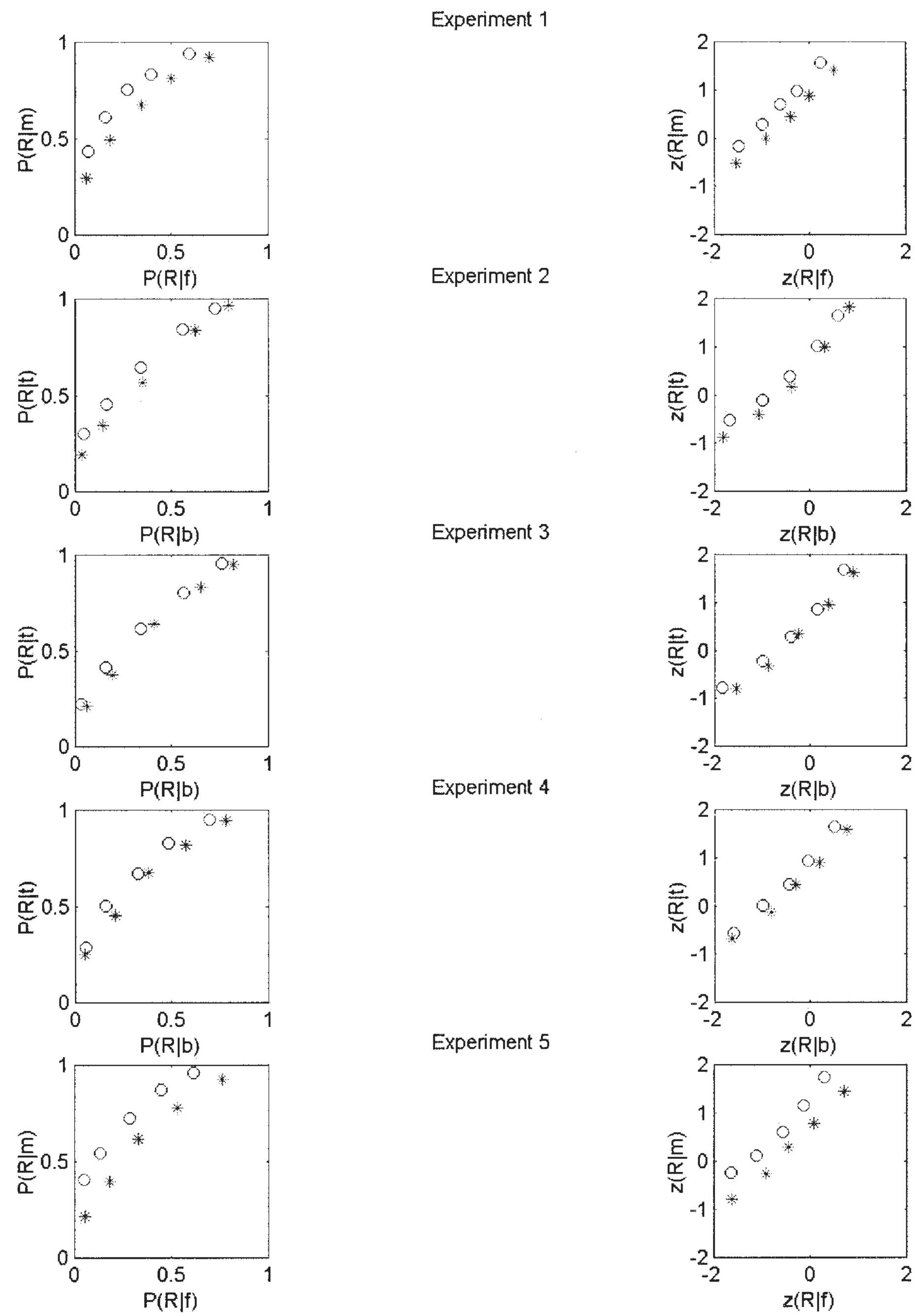

Figure 3. Source-recognition receiver operating characteristics (ROCs) and receiver operating characteristics plotted in standard scores (z-ROCs) for Experiments 1–5. The circles represent the superior recognition condition in each experiment, for example, the repeated condition in Experiment 1, and the asterisks represent the inferior recognition condition, for example, the single-presentation condition. R|m = rating, given source is male voice; R|f = rating, given source is female voice; R|t = rating, given source is top; R|b = rating, given source is bottom.

group z-ROC. For this experiment, the R^2s, listed in Appendix B, are 0.995 for the single and 0.997 for the repeated presentation. This further confirms the point made in Hilford et al. (2002) that the departure from linearity seen in the preceding source-recognition experiments and in those that follow is very small, although full analysis of the z-ROCs shows that the small departure to concavity is statistically significant. Here the departure accounts for less than 1% of the variance in the z-ROC. The R^2s for Experiments 2, 3, 4, and 5, also listed in Appendix B, have similar values.

Individual data. For a full statistical evaluation of the ROCs and z-ROCs, we translated each individual's ratings into ROCs and corresponding z-ROCs. Each ROC was fitted with a quadratic equation to obtain a measure of convexity, the quadratic constant. Each z-ROC was fitted with a linear equation. These fits furnished slopes and intercepts, and the latter were used as an index of accuracy, d'_2. Each z-ROC was also fitted with quadratic equations to obtain quadratic constants that measured departure from rectilinearity. The means of these statistics—ROC quadratic constants, z-ROC linear slopes, and z-ROC quadratic constants are shown in Table 1. The mean linear intercepts for the z-ROCs are shown in Table 2.

Although most of the 70 participants used most or all of the range of confidence ratings, 10 used fewer than three of the ratings in either the single or repeated presentations. For example, a participant would use only two of the rating categories for one of the conditions. This is understandable when participants are dealing with four different distributions arranged along a single dimension. We could not compute quadratic constants for these participants. Therefore, tests of the quadratic constant are based on 60 participants' data. Examination of the other statistics obtained from this reduced group, for example, linear slopes and the intercepts of the z-ROCs, were close to those for the full group of 70.

In tests of the regularities that follow, if two related statistics differ, for example, the ROC quadratics for the single and repeated conditions, then tests on those statistics, for example, whether they differ from zero, were carried out separately on each. If they do not differ, then a single test has been carried out on the mean of the two.

Tests of the regularities find the following:

Table 1
Mean Source Recognition ROC and z-ROC Statistics: Experiments 1–5

Experiment and variable	ROC Quadratic constant *M*	ROC Quadratic constant *SE*	z-ROC Linear slope *M*	z-ROC Linear slope *SE*	z-ROC Quadratic constant *M*	z-ROC Quadratic constant *SE*
1. Repetition						
Single	−1.10	.23	0.96	.04	0.22	.06
Repeated	−2.74	.43	0.97	.04	0.04	.09
2. List Length						
Long	−0.42	.12	0.98	.03	0.25	.04
Short	−0.96	.28	1.01	.04	0.28	.05
3. Encoding						
Shallow	−0.61	.13	1.00	.05	0.10	.05
Deep	−0.75	.22	1.00	.05	0.18	.07
4. Word Frequency						
High	−0.80	.15	1.01	.04	0.17	.04
Low	−1.06	.20	1.00	.03	0.18	.06
5. Aging						
Old	−1.29	.38	0.96	.05	0.09	.04
Young	−1.19	.33	0.99	.04	0.34	.09

Note. ROC = receiver operating characteristic; z-ROC = receiver operating characteristic plotted in standard scores.

Table 2
Linkage of Experimental Conditions: Item and Source Accuracy Measures in Experiments 1–5

Experiment and variable	Item d'	Source d'_2
1. Repetition		
Single	0.95	1.10
Repeated	1.32	1.47
2. List Length		
Long	1.31	0.82
Short	2.07	1.08
3. Encoding		
Shallow	3.62[a]	0.65
Deep	3.86[a]	0.89
4. Word Frequency		
High	0.57	0.88
Low	1.17	1.18
5. Aging		
Old	0.91[b]	0.89
Young	1.40[b]	1.49

[a] Encoding item recognition d's are high because the items tested in the source recognition test were repeated as old items in the item recognition test. [b] The aging item recognition accuracy measure is d'_2.

1. Convex ROCs. The two ROCs' mean quadratics were negative. Test of the two mean constants' difference was statistically significant, $F(1, 59) = 8.87$, $MSE = 18.28$, $p = .004$. Departure of each from zero was statistically significant: for single, $t(59) = 4.77$, $SE = 0.23$, $p < .001$; for repeated, $t(59) = 6.34$, $SE = 0.43$, $p < .001$.

2. Concave z-ROCs. The two z-ROCs' quadratic means, both positive, did not differ at a statistically significant level, $F(1, 59) = 2.04$, $MSE = 0.94$, $p = .158$. Test of their mean found a statistically significant departure from zero, $t(59) = 2.40$, $SE = 0.05$, $p < .02$.

3. Constant slopes. The linear slopes of the two z-ROCs did not differ reliably, $F(1, 69) = 0.12$, $MSE = 0.04$, $p = .727$.

4. Slopes of 1.00. The mean slope of the z-ROCs did not differ reliably from 1, $t(69) = 1.42$, $SE = 0.03$, $p < .20$.

5. Mirror effect. We tested for presence of a mirror pattern using the confidence ratings. These contained the maximum amount of information on the underlying distributions' locations. The mean ratings for the four conditions presented in this experiment are shown in the first line of Table 3. If a mirror pattern held, then the mean ratings would be ordered as follows: Male Voice 2 (m_2), Male Voice 1 (m_1), Female Voice 1 (f_1), Female Voice 2 (f_2), where 1 and 2 refer to single and repeated presentation: $R(m_2) < R(m_1) < R(f_1) < R(f_2)$. The order does appear. When the two sources and the two strength levels are organized into a 2×2 table, the mirror effect is then seen as a crossover interaction. This interaction is statistically significant, $F(1, 69) = 28.30$, $MSE = 0.18$, $p < .001$.

6. Linked recognition accuracy. Item-recognition d's (see

Table 3
Source Mirror Effect: Mean Confidence Ratings for Experiments 1–4

Experiment and variable	Source A		Source B		
	A_s	A_w	B_w	B_s	*SE*
1. Repetition	2.47	2.77	4.20	4.44	.07
2. List length	2.80	3.09	4.04	4.16	.05
3. Encoding	2.97	3.01	3.88	4.12	.06
4. Word frequency	2.80	2.90	4.00	4.27	.06

Note. Source A indicates items read in a male voice, and Source B indicates items read in a female voice in Experiments 1 and 3. Source A indicates top position, and Source B indicates bottom position in Experiments 2 and 4. The subscripts S and W refer to the strong and weak, conditions in each experiment; for example, S indicates repeated and W indicates single presentation in Experiment 1.

Table 2) based on hit and false alarms were reliably different, $F(1, 69) = 86.84$, $MSE = 0.12$, $p < .001$, as would be expected. The corresponding source-recognition accuracy measures, d'_2, were reliably different, $F(1, 69) = 19.73$, $MSE = 0.46$, $p < .001$. The other indication of linkage is the correlation of item accuracy (d') and source accuracy (d'_2) within each experimental condition. The correlation in the single condition was 0.69, and in the repeated condition, it was 0.77. Both are statistically significant.

To summarize, the results replicate the three regularities found in preceding work (Hilford et al., 2002). The ROCs were convex, the z-ROCs were slightly concave, and the linear slopes of the z-ROCs were near 1.00. The results, in addition, support three additional regularities. The slopes of the z-ROCs remained constant. The source-recognition confidence ratings means formed a mirror pattern. Accuracy in the two tasks was linked. A variable that affected item recognition, here repetition, had a similar effect on source recognition. Within each experimental condition, individuals' performance on the two tasks was linked, in other words, correlated.

We noted earlier that we constructed the tasks so that the two sources, here male and female voices, were matched. We can use the item d's to check their equivalence. The item d' for the male voice was 1.14; for female, it was 1.13. The two are not reliably different, $F(1, 69) = 0.43$, $MSE = 0.05$, $p = .515$. There was no interaction with repetition conditions, $F(1, 69) = 1.82$, $MSE = 0.07$, $p = .182$.

Experiment 2: List Length

Another variable that reliably affects item-recognition accuracy is length of the study list (Bowles & Glanzer, 1983; Gronlund & Elam, 1994; Ratcliff et al., 1994; Strong, 1912; Yonelinas, 1994). Short lists produce better item recognition than long lists. We now evaluate the effect of this variable on source recognition.

Method

Participants studied two lists, one of which was short and the other of which was long. After each study list, they had a source-recognition test followed by an item-recognition test. The source test consisted solely of studied words. For each word, participants indicated whether the word appeared, during study, on the top or bottom of the screen, using a six-level confidence rating. Following the source test, participants were given an item-recognition test consisting of 10 old words (seen on the study list) and 10 new words. In the item-recognition test, the participants chose each word as either old, having been in the study list, or new, not having been in the study list.

Participants. Participants were 56 undergraduates.

Materials. Three hundred and sixty words chosen from Paivio et al. (1968) served as the list from which the participants' study and test lists were created. From this list, 180 words were randomly selected for the long list, and 50 words were randomly selected for the short list. Additional words were selected from the remainder of the list to be used as practice list items, as filler words for the study and test lists, and as new items for the item-recognition tests.

Procedure. The experiment began with practice lists to familiarize participants with the procedure. The practice procedure was the same as that for the main lists except for the number of words used. The participants first viewed a six-word study list in which three of the words appeared on the top and three of the words appeared on the bottom of the screen. This list was followed by a source-recognition test and then an item-recognition test. The practice source-recognition test consisted of three of the studied words for which the participants indicated source: top or bottom. The practice item-recognition test consisted of the other three studied words (old) and three new words.

The full experiment consisted of two blocks. The procedure in each block was the same. The only difference was the length of the study list. Half the participants had the short 50-word list first. The other half had the long 180-word list first. The presentation location (top or bottom of the screen) of the study-list items was selected at random for each word, with the restriction that half were assigned to each position. The words were presented in uppercase letters for 1,000 ms. A 750-ms interstimulus interval separated each word. Participants were told to study the words carefully and that they would be tested on them.

The source-recognition test followed the study list with the test items shown in the center of the computer screen. The source test list for both long and short lists consisted of the last 40 study words (before the final 5 filler words) rerandomized so that test order differed from the study list. This was done so that any effects of list length would not be confounded with delay effects. The participants used a six-point confidence rating scale in the source-recognition test as in Experiment 1.

Following each source-recognition test, participants took a yes–no item-recognition test. The recognition test consisted of 10 old words and 10 new words. The 10 old words were those that directly preceded the 40 study items used for the source-recognition test. Both the long and short study lists had, in addition to the words mentioned above, filler items. The study lists started and ended with five untested filler items. The source-recognition tests started with five unscored filler items. The item-recognition tests started with two unscored filler items.

Results

Group data. The overall results based on group ROCs and z-ROCs are shown in the panels of the second row of Figure 3. The data show the same pattern as those for Experiment 1: The ROCs are convex, the z-ROCs are concave, the linear slopes of the z-ROCs are near 1.00, and the two z-ROCs are separated. The linear R^2s are 0.969 and 0.972. The summary statistics supporting these statements are presented in Appendix B.

Individual data. Using the same analyses as those for Experiment 1 on the individual data, we produced the mean statistics shown in Tables 1, 2, and 3. The same regularities found in Experiment 1 hold in Experiment 2.

1. Convex ROCs. The quadratic constants for the two ROCs do not differ reliably, $F(1, 55) = 2.52$, $MSE = 6.60$, $p = .118$. The combined mean, -0.68, differs reliably from zero, $t(55) = 4.48$, $SE = 0.15$, $p < .001$.

2. Concave z-ROCs. The two z-ROCs' quadratic constants do not differ reliably, $F(1, 55) = 0.32$, $MSE = 0.06$, $p = .575$. Their combined mean, 0.26, differs reliably from zero, $t(55) = 8.69$, $SE = 0.030$, $p < .001$.

3. Constant slopes. The slopes of the z-ROCs for the two conditions do not differ reliably, $F(1, 55) = 2.54$, $MSE = 0.02$, $p = .117$.

4. Slopes of 1.00. The mean slope of the z-ROCs does not differ reliably from 1.00, $t(55) = 0.11$, $SE = 0.02$, $p < .80$.

5. Mirror effect. The mean ratings show the mirror pattern. See Table 3. The Source × Length interaction is statistically significant, $F(1, 55) = 6.11$, $MSE = 0.38$, $p = .017$.

6. Linked recognition accuracy. Item-recognition accuracy differs reliably, $F(,1\ 55) = 22.54$, $MSE = 1.42$, $p < .001$, as would be expected. The corresponding source-recognition accuracy also differs reliably, $F(1, 55) = 4.36$, $MSE = 0.87$, $p = .041$. The correlations between item and source recognition, shown in Table 4, are statistically significant.

All six regularities are supported.

A check of the item d's shows that the sources were matched. The top position item d' was 1.72, and the bottom was 1.66. The difference is not statistically significant, $F(1, 55) = 0.31$, $MSE = 0.58$, $p = .578$, with no evidence of interaction with list length, $F(1, 55) = 0.61$, $MSE = 0.86$, $p = .438$.

Table 4
Linkage of Individuals' Performance: Correlations of Participants' Item and Source Accuracy Within Each Experimental Condition in Experiments 1–5

Experiment and variable	Correlations
1. Repetition	
Single	.69*
Repeated	.77*
2. List length	
Long	.28*
Short	.39*
3. Encoding	
Shallow	.38*
Deep	.17
4. Word frequency	
High	.66*
Low	.51*
5. Aging	
Old	.37
Young	.60*

* $p \le .05$.

Experiment 3: Encoding

Another well-established effect on item-recognition accuracy is that of variation in encoding, variation in the processing that participants carry out on items in the study list. Tasks that require the participant to become more fully engaged in processing the item, for example, having them classify the words in some way (low frequency vs. high frequency), will produce more accurate item recognition than tasks that have the participants simply view the word or note a superficial characteristic, for example, whether the word contains a particular letter (Craik & Tulving, 1975; Glanzer & Adams, 1990; Glanzer, Kim, et al., 1999; Hilford et al., 1997). The first type of encoding tasks is often labeled *deep*; the second is often labeled *shallow*.

The effect of encoding on source recognition is of particular interest in relating source recognition to item recognition. The 2D-SDT model, as interpreted here, predicts that the effect of encoding tasks on source- and item-recognition accuracy should covary. Some investigators have presented evidence to the contrary: that encoding variation has opposed effects on source-recognition accuracy and item-recognition accuracy (Lindsay & Johnson, 1991). The experiment used to support that view, however, has special characteristics that will be discussed in detail later in this article.

Method

Participants heard a list of study words, half of which were read in a man's voice and half of which were read in a woman's voice. For half of the words heard in each voice, the participants carried out a less effective, shallow encoding task. For the other half, they carried out a more effective, deep encoding task. Two successive tests followed the completion of the study list, a source-recognition test and an item-recognition test. For the source-recognition test, participants viewed a list of all the old words, the words presented in the study list, so designated to the participants. For each word, participants indicated whether the word was spoken during study by the man or by the woman, using a six-level confidence rating system. Following the source list, the participants completed an item-recognition test.

Participants. Participants were 36 undergraduates.

Materials. We used the 720 digitally recorded words used in Experiment 1 to construct the individually randomized main lists for the experiment. The 45 additional recorded words furnished the practice and filler items. The study list began with 5 untested filler items and ended with 5 untested filler items drawn from those 45 words.

Procedure. The experiment began with a short practice study list and test list using a procedure that was the same as that for the main list: a study list followed by a source-recognition test and then an item-recognition test. The practice study list consisted of 10 words, as did the following practice source-recognition test. The practice item-recognition test contained 5 old words and 5 new words.

The study-list procedure was similar to that used in Experiment 1. For each participant, a study list of 90 words was selected at random from the full list of 360 words, randomly ordered. The voicing (male or female) of the items was also selected at random for each word with the restriction that half were in the male voice and 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 they would be tested on them. While the words were presented through the headphones, the computer screen was blank. After each word, one of two different encoding tasks was presented on the screen. The more effective encoding task, deep, was a frequency judgment in which participants were asked to indicate whether the presented word was a common or uncommon word. The less effective encoding task, shallow, was a letter identification task. For this, participants were asked to identify whether the just-presented word included either the letter *K* or the letter *I*. (For the first 5 participants, the letters were *L* or *A*. These alternatives produced slightly more than 50% "yes" responses. We therefore switched to the letters *K* or *I*, which produced closer to 50% "yes" responses.) Participants indicated their selection in each condition on one of two pairs of keys on their keyboard. A 1,250-ms interstimulus silent interval separated successive words. During this interval, a row of asterisks appeared on the computer monitor.

The source-recognition test, presented visually on the computer screen, followed the study list directly. The source test lists consisted of the 90 study words from the study list rerandomized. All of the test items appeared in the middle of the screen in uppercase letters. Participants viewed the words one at a time and, using the six-point confidence rating procedure described in the preceding experiments, indicated whether each word was spoken by the man or the woman.

Following the source-recognition test, participants took a yes–no item-recognition test. Participants were presented with a list of 40 words, 20 old and 20 new, and asked to indicate for each word whether the item was old or new. The words were presented on the computer screen centered and in uppercase letters. The old words in this test had been seen twice, once in the study list and once in the source test. The repetitions furthermore had the advantages of widely spaced repetition and a very short test list. These advantages produced the very high item d's in Table 2.

Results

Group data. The group ROCs and z-ROCs are shown in the third row of panels in Figure 3. The group data show the earlier regularities: convex ROCs, z-ROCs slope near 1.0, and a slight concavity of z-ROCs. They also show a difference in both the ROCs and the z-ROCs indicating a difference in accuracy. The linear R^2s are 0.990 and 0.971. See Appendix B for summary statistics.

Individual data. The means of the individuals' ROC statistics are shown in Tables 1 and 2, and the mean ratings are shown in Table 3. All the effects found in the preceding experiments appear in Experiment 3.

1. Convex ROCs. The two ROCs' negative quadratic constants do not differ reliably, $F(1, 35) = 0.24$, $MSE = 2.89$, $p = .630$. Their mean differs reliably from zero, $t(35) = 5.27$, $SE = 0.13$, $p < .001$.

2. Concave z-ROCs. The two z-ROCs' positive quadratic constants do not differ reliably, $F(1, 35) = 0.81$, $MSE = 0.27$, $p = .374$. Their mean differs reliably from zero $t(35) = 3.40$, $SE = 0.04$, $p < .002$.

3. Constant slopes. The slopes for the two conditions do not differ, $F(1, 35) = 0.00$, $MSE = 0.03$, $p = .965$.

4. Slopes of 1. The means of the slopes of the two z-ROCs do not differ reliably from 1.00, $t(35) = 0.11$, $SE = 0.03$, $p < .80$. It could be argued that the slope near 1.00 in the shallow encoding condition is the result of chance performance. The performance is not, however, chance. The mean d'_2 for the shallow encoding condition, 0.65, is different from zero, $t(35) = 8.76$, $SE = 0.07$, $p < .001$, at a statistically significant level. Given the fact that none of the three source-recognition slopes in Hilford et al. (2002) or the other nine slopes in this study differ from 1.00, all with d'_2s reliably greater than zero and ranging up to 1.49, there is reason to consider the slope of 1.00 in the shallow encoding condition as another example of the slope regularity.

5. Mirror effect. There is a reliable mirror effect in the mean confidence ratings. The Source $\times$ Encoding interaction is $F(1, 35) = 7.95$, $MSE = 0.09$, $p = .008$.

6. Linked recognition accuracy. The two encoding conditions differ in item-recognition accuracy, $F(1, 35) = 7.99$, $MSE = 0.26$, $p = .008$. The two encoding conditions also differ reliably in source-recognition accuracy, $F(1,35) = 7.57$, $MSE = 0.29$, $p = .008$. Within each condition, the correlation between item and source accuracy is positive and is at a statistically significant level in the shallow encoding condition.

Again, the six regularities were replicated. Check of the item d's indicates that the two sources within each experimental condition can be considered matched. The d's are not reliably different with 3.82 for the male voice and 3.67 for the female voice. The interaction with encoding condition is also not statistically significant.

Experiment 4: Word Frequency

Normative word frequency is another variable that has an established effect on item-recognition accuracy (Allen & Garton, 1968; Balota & Neely, 1980; Bowles & Glanzer, 1983; Garton & Allen, 1968; Glanzer & Adams, 1990; Glanzer & Bowles, 1976; Gorman, 1961; Gregg, 1976; Hilford et al., 1997; Kim & Glanzer, 1993; Kinsbourne & George, 1974; McCormack & Swenson, 1972). Low-frequency words are more accurately recognized than are high-frequency words in item recognition. We now examine the effect of word frequency on source recognition.

Method

Participants viewed a randomly mixed list of 70 high- and 70 low-frequency words. Half of each frequency set appeared on the top of the computer monitor, and half appeared on the bottom of the monitor. Following the end of the study list, participants were given a recognition test list that included the 140 study words plus 140 new words, half of which were high- and half of which were low-frequency words. Participants were asked whether each word was old (i.e., had appeared during study) or new.

For each word responded to as "old," participants indicated whether the word had appeared on the top or the bottom of the screen, using a confidence rating scale.

Participants. Participants were 51 undergraduate students.

Materials. We selected 310 words from Kučera and Francis (1967). Of these words, half (155) were high-frequency words with a mean normative frequency (per million) of 289.59, and half were low-frequency words with a mean normative frequency of 0.95. The list of 280 study and test words (plus 10 filler words) was randomly selected for each participant from this set for the main study list.

Procedure. The experiment began with a short practice session in which a study list of 10 words was presented followed by an item and source test list of the 10 studied words, plus 10 new words in mixed order.

For each participant, a study list of 70 high-frequency words and 70 low-frequency words was selected from the full list of 310 words and randomly ordered. Five untested filler items were added to the beginning and five to the end of this main list. The list was presented on a computer monitor, with each word in uppercase letters, either on the top or bottom of the screen, for a duration of 1,000 ms. A 750-ms interstimulus interval separated successive words. Location was randomly selected for each word with half of the words of each frequency in each position. Participants were told to study the words and that they would be tested on them.

A randomized item- and source-recognition test followed the study list directly, presented visually on the computer monitor. All of the test items appeared in the middle of the screen in uppercase letters. Each test list consisted of the 140 study-list words randomly mixed with an equal number of new words. The testing procedure was the same as that in Experiment 1. Participants viewed the words one at a time and for each word, indicated whether it was old, had appeared during study, or was new. Participants entered their choices using either a key marked *O* for "old" or a key marked *N* for "new." If a participant selected *N*, the next word in the test list was presented for old–new identification. If a participant selected *O*, he or she was asked whether the word had appeared on the top or bottom of the screen using a six-point confidence rating.

Results

Group data. The group source-recognition data are shown in the panels of the fourth row of Figure 3. The two ROCs are convex. The two z-ROCs are slightly concave with linear slopes near 1.00. Both pairs are separate. The linear R^2s are 0.986 and 0.990. See Appendix B for summary statistics.

Individual data. The means of the individuals' ROC and z-ROC statistics are shown in Tables 1 and 2, and the mean ratings are shown in Table 3. Four participants did not use a sufficient number of rating categories for either high-frequency or low-frequency words to permit application of quadratic equations to their data. Those data are therefore not represented in the evaluation of the convexity of the ROCs and the concavity of the z-ROCs, reducing the degrees of freedom in analysis of the quadratics. Analysis of the mean statistics shows the same effects as the preceding experiments.

1. Convex ROCs. The two ROCs' negative quadratics do not differ reliably, $F(1, 46) = 0.77$, $MSE = 3.82$, $p = .385$. Their mean does differ reliably from zero, $t(46) = 7.51$, $SE = 0.12$, $p < .001$.

2. Concave z-ROCs. The two z-ROCs have positive quadratic constants that do not differ reliably, $F(1, 46) = 0.04$, $MSE = 0.26$, $p = .839$. Their mean differs from zero, $t(46) = 5.12$, $SE = 0.03$, $p < .001$.

3. Constant slopes. The slopes for the two conditions do not differ, $F(1, 50) = 0.47$, $MSE = 0.03$, $p = .495$.

4. Slopes of 1.00. The mean slopes of the z-ROCs do not differ reliably from 1.00, $t(50) = 0.22$, $SE = 0.03$, $p < .80$.

5. Mirror effect. There is a reliable mirror effect in the mean confidence ratings. The Source $\times$ Word Frequency crossover interaction is $F(1, 50) = 8.33$, $MSE = 0.20$, $p = .006$.

6. Linked recognition accuracy. The two word frequency conditions differ reliably in source-recognition accuracy, as measured by d'_2, $F(1, 50) = 7.56$, $MSE = 0.59$, $p = .008$; and in item-recognition accuracy $F(1, 50) = 113.28$, $MSE = 0.16$, $p < .001$. The correlations of item- and source-recognition accuracy within each experimental condition are shown in Table 4. Both are statistically significant.

Again, all the regularities are replicated. A check of the item-recognition d's indicates that the two sources are matched within each experimental condition. The d' for the top is 0.88, and the d' for the bottom is 0.86, $F(1, 50) = 0.27$, $MSE = 0.04$, $p = .609$. There is no evidence of interaction with word frequency, $F(1, 50) = 0.30$, $MSE = 0.03$, $p = .586$.

We also analyzed the item-recognition data for mirror effect. (This is the first experiment in the set reported in this article in which the variable applies to both new and old items. Here the new items can be either high frequency or low frequency. In the preceding experiments, new items cannot be distinguished in this manner, for example, as single or repeated, or from a short or long list.) We analyzed the proportion of "yes" responses because participants did not rate confidence in item recognition. The mirror pattern in this case is

$$\mathrm{P(SN)} < \mathrm{P(WN)} < \mathrm{P(WO)} < \mathrm{P(SO)}$$

where P is the proportion of "yes" responses, N is new, O is old, W is weak, and S is strong. Because weak is high frequency, and strong is low frequency, this can be written as

$$\mathrm{P(LN)} < \mathrm{P(HN)} < \mathrm{P(HO)} < \mathrm{P(LO)}$$

where L is low frequency, and H is high frequency. The corresponding means found were the following:

$$0.13 < 0.18 < 0.42 < 0.59.$$

An analysis of variance found a reliable Frequency $\times$ Old versus New interaction, $F(1, 50) = 142.97$, $MSE = 0.004$, $p < .001$.

Experiment 5: Aging

The linkage of item- and source-recognition accuracy with aging as the variable is fully established. Aging has been demonstrated to affect item recognition (Clark, 1980; Dorfman, Glanzer, & Kaufman, 1986; Erber, 1974, 1978; Gordon & Clark, 1974; Harkins, Chapman, & Eisdorfer, 1979; Kausler & Kleim, 1978). It has also been demonstrated to affect source recognition (Hashtroudi et al., 1989; Henkel, Johnson, & DeLeonardis, 1998; Johnson, DeLeonardis, Hashtroudi, & Ferguson, 1995; Schacter et al., 1991). The linkage is so much an established fact that major effort has been focused on whether aging depresses source recognition more than it does item recognition. See the meta-analysis of 46 experiments by Spencer and Raz (1995). We now investigate whether the other regularities hold as well for both young and old participants. We also complete the lining up of source-recognition regularities with item-recognition regularities by obtaining item-recognition ROCs and z-ROCs.

Method

Two groups of participants, one old and one young, were tested on both source-recognition memory and item-recognition memory. The two sources were a male voice and a female voice. We obtained confidence ratings for both the item- and source-recognition tests. These two sets of ratings permit the plotting and analysis of both item- and source-recognition ROCs and z-ROCs. This experiment differs from the preceding experiments in that the experimental variable, age, is a between-group variable.

Participants. There were two groups of participants: old and young. The old group consisted of 24 individuals who were 65 years old or older (M = 73.37, SD = 5.21). Their vision and hearing were normal or corrected to normal. Their educational degrees were the following: high school, 2; associate, 1; bachelor's, 11; master's, 9; doctorate, 1. An additional individual, age 85, gave chance responses on his test. His data are not included. The old participants were paid for the session. The young group consisted of 42 undergraduate students, mean age 18.77 years (SD = 1.84).

Materials. The material was the same as in Experiment 1. Study and test lists were individually randomized as in the preceding experiments.

Procedure. Half of the participants in each group took the item-memory study and test first, and the other half took the source-memory study and test first. At the beginning of both tests, participants were told that their memory for both source and item memory would be tested. During both study lists, participants heard 56 words through earphones connected to the computer. Half of the words were presented in a male voice. The other half were presented in a female voice. The first four and the last four items were unscored filler items.

During the source-memory test, 48 old words were presented visually; during the item-memory test, 48 studied words and 48 new words were presented visually. Participants used a six-point confidence rating scale on both tests with instructions for the rating scale displayed on the computer monitor during testing.

Results

Group data. The group source ROCs and z-ROCs are shown in the bottom panels of Figure 3. They show the same characteristics as the preceding experiments. The item-recognition ROCs and z-ROCs are shown in Figure 4.

Source individual data. The mean statistics for the individuals' ROCs and z-ROCs are shown in Tables 1, 2, and 3. Three participants in the young group did not use a sufficient number of ratings to permit computation of quadratic fits. This holds also for their item-recognition data. Mirror effects are not analyzed here because the variable is a between-groups variable. The standard arrangement for evaluating mirror effects is with a within-groups variable, as in the preceding experiments. The remaining five source-recognition regularities hold.

1. Convex ROCs. The two groups' negative quadratics are not reliably different, $F(1, 61) = 0.02$, $MSE = 12.47$, $p = .880$. Their mean, -1.23, is reliably different from zero, $t(62) = 4.88$, $SE = 0.25$, $p < .001$.

2. Concave z-ROCs. The two groups' positive mean quadratic constants do not differ reliably, $F(1, 61) = 3.14$, $MSE = 0.60$, $p = .081$. Their mean differs reliably from zero, $t(62) = 4.19$, $SE = 0.06$, $p < .001$.

3. Constant slopes. The two groups' slopes do not differ reliably, $F(1, 64) = 0.51$, $MSE = 0.05$, $p = .479$.

4. Slopes of 1.00. The mean z-ROC slope does not differ reliably from 1.00, $t(65) = 0.72$, $SE = 0.03$, $p < .80$.

5. Linked recognition accuracy. The two groups' source-recognition accuracy means, d'_2, differ at a statistically significant level, $F(1, 64) = 8.55$, $MSE = 1.29$, $p = .005$. Item-recognition accuracy also differs reliably. The correlations between item and source recognition are both

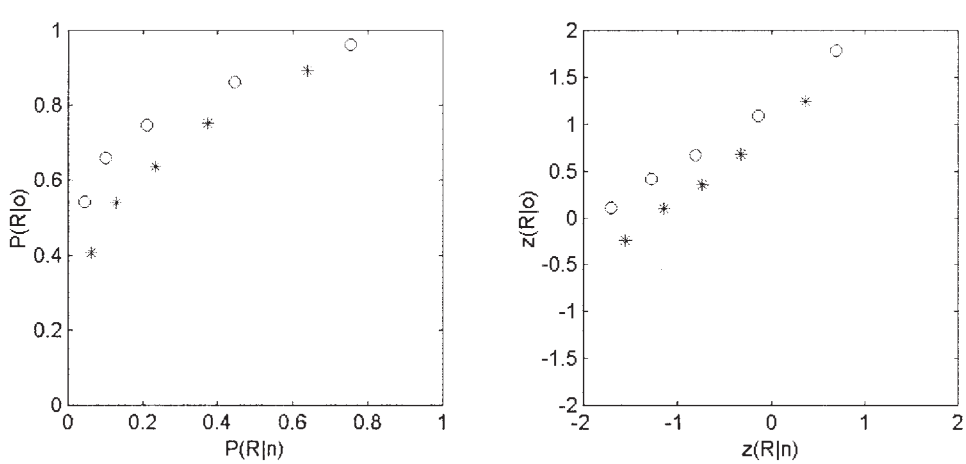

Figure 4. Item-recognition receiver operating characteristics (ROCs) and receiver operating characteristics plotted in standard scores (z-ROCs) for Experiment 5. The circles represent the young group, and the asterisks represent the old group. R|o = rating, given item is old; R|n = rating, given item is new.

positive and statistically significant in the young group. See Table 4.

Item group data. We have ROCs and z-ROCs for item recognition as well as source recognition in this experiment. The group item-recognition data are displayed in Figure 4. The group ROCs are convex and separated. The group z-ROCs are linear and separated. The summary statistics are given in Appendix B.

Item individual data. Analysis of the mean statistics derived from the individuals' ROCs and z-ROCs show the established item-recognition regularities to which we referred earlier.

1. Convex ROCs. The groups' negative quadratic constants do not differ reliably, $F(1, 61) = 1.18$, $MSE = 12.23$, $p = .282$. Their mean shows reliable convexity, $t(62) = 3.53$, $SE = 0.55$, $p < .001$. This is a standard finding for item recognition.

2. Linear z-ROCs. The two groups did differ reliably in the quadratic constants of their z-ROCs, $F(1, 61) = 5.17$, $MSE = 0.61$, $p = .026$. Each mean, however, did not differ reliably from zero: young, 0.08, $t(39) = 1.57$, $SE = 0.13$, $p < .20$; old, $-.025$, $t(24) = 1.98$, $SE = 0.13$, $p < .10$.

3. Slopes less than 1.00. Both groups show z-ROC slopes that are less than 1.00: young, 0.68, $t(41) = 11.30$, $SE = 0.03$, $p < .001$; old, 0.82, $t(23) = 3.65$, $SE = 0.05$, $p < .002$. This is a standard finding for item recognition (Glanzer, Kim, et al., 1999; Ratcliff et al., 1994, 1992).

4. Varying slopes. The slopes vary across the two groups as a function of accuracy. The difference is statistically significant, $F(1, 64) = 5.97$, $MSE = 0.11$, $p = .017$. This again is a standard finding for item recognition (Glanzer, Kim, et al., 1999).

A check of the equality of item recognition for the two sources finds that it holds for the young group, $t(41) = 0.32$, $SE = 0.08$, $p < .60$; but not for the old group, $t(23) = 3.60$, $p < .02$. The old group was more accurate on words from the female voice ($d'_2 = 1.11$) than from the male voice ($d'_2 = 0.71$). This inequality does not, however, disrupt the linkage between source and item accuracy. The young group shows greater accuracy in both source and item accuracy.

Discussion

The five experiments support six regularities for source recognition.

1. Convex ROCs
2. Slightly concave z-ROCs
3. z-ROCs with linear slopes of 1.00
4. Constancy of z-ROC slopes
5. Mirror effect
6. Linked accuracy in item and source recognition

The findings on the first three regularities replicate previous findings (Hilford et al., 2002), and they add the information that they hold as accuracy levels are changed (see Regularity 4). It could be argued that the fifth regularity, the mirror effect, does not require any special explanation or model. It could be argued that it is common sense that the participant simply responds with extreme ratings for words that are well recognized and middling ratings for words that are not. It is also common sense that apples fall. But common sense alone does not explain a regularity. Moreover, the regularity of the mirror effect in source recognition parallels a regularity in item recognition. And that regularity is not handled by common sense. There is no commonsense reason why new items (low vs. high frequency) should be responded to differently when old (low vs. high frequency) items are responded to differently in item-recognition tests. An explanation, in other words, model, is required to explain both item and source mirror effects. The model offered here covers the parallels between the item- and source-recognition regularities.

The sixth regularity, the linkage of item and source recognition, is supported in all five experiments reported here. It is further supported by findings of Dodson, Holland, and Shimamura (1998). They reported two experiments. In Experiment 1, the similarity of the sources was varied. In Experiment 2, attention during test was varied. In both experiments, the condition that reduced item-recognition accuracy also reduced source-recognition accuracy. See their values for specific source identification in their Tables 3 and 6. The correlated reductions of item recognition can be seen by computing d's for old–new recognition and false alarm rates in those tables.

Source Regularities 1 (convex ROCs) and 6 (mirror effects) are the same as the effects in item recognition. Source Regularity 2, the slightly concave z-ROCs, differs. It can, however, be handled in an extended 2D-SDT framework (see Hilford et al., 2002). Source Regularities 3 and 4 (linear z-ROC slopes of 1.00 that remain constant across accuracy changes) can be handled within 2D-SDT. The two sources at each level of the experimental variable within an experiment are matched and have the same history. They do not differ as do a new and old distribution in an item-recognition experiment. There the items of the old distribution are changed by exposure in a study list, whereas items of the new distribution are not so exposed. The two sources at each level of the experimental variables (e.g., repetition) are matched in history, and both have been changed by the same type of exposure.

It could be argued that relations found between item- and source-recognition accuracy with repetition (Experiment 1) and with change in list length are not surprising. Repetition gives the participants two opportunities to collect both item and source information. Long lists impair the retrievability of both stored item and source information. This argument cannot, however, be made for encoding effects, as shown in Experiment 3. The encoding operations were applied solely to the words presented. They make no reference to the sources. In fact, if the better encoding operation, one that improves item recognition, produces differences in attention to the words, then it would be expected that the more effective operation for item recognition should draw attention away from source information and produce a drop in source recognition.

The same reasoning applies even more strongly to the effect of word frequency. There is no reason, outside of the 2D-SDT theory, why low-frequency words should produce greater source-recognition accuracy than high-frequency words. Again, if some assumption about attention difference is made, then word frequency effects on source recognition should be the reverse of their effects on item-recognition accuracy.

There is considerable variation in the correlations between item- and source-recognition accuracy seen in Table 4. The variations stem from the fact that the experiments differ considerably in *N*, the number of participants, the variables used, and the accuracy induced by these variables. For example, the correlation between *N* and the size of the 10 correlations is .58 ($p = .08$).

The importance of linkage for a theory that includes both item and source recognition is implicitly recognized in attempts to find dissociations between the two. Before discussing the existence of such dissociations, we note first that, if they were to exist, they could be handled by an extended 2D-SDT model that includes an attention component, the mixture model presented by Hilford et al. (2002). In that extension, the attention component is brought in to explain the slight concavity of source-recognition z-ROCs. It is not necessary, however, to invoke that component at this point.

There are three studies that present sets of data that show interactions that might be considered contrary to the general 2D-SDT model: those of Glisky, Polster, and Routhieaux (1995), Jurica and Shimamura (1999), and Lindsay and Johnson (1991).

The Glisky et al. (1995) and the Lindsay and Johnson (1991) studies are not problematic for simple 2D-SDT as presented here when examined fully. The Jurica and Shimamura (1999) results may present a problem, which can, however, be handled with 2D-SDT as extended in Hilford et al. (2002).

Glisky et al. (1995) separated a group of 48 old participants twice. One separation was into high-frontal and low-frontal (capacity) groups. The other separation was into high medial-temporal and low medial-temporal (capacity) groups. The partitioning was done on the basis of a battery of test scores. They then compared the forced-choice scores for item and source recognition in each of the four groups. The results are presented in Table 5. Glisky et al. focus on the small difference in item recognition in the frontal groups as opposed to the large difference in the medial temporal groups and the reversed pattern in the source-recognition data. The finding is an important one for the determination of the way that different areas of the brain function in these tasks. The findings, however, are not problematic for 2D-SDT. They indicate that the classifications are related to differences in the participants' efficiency in handling different aspects of information. Pairs of triangles, like those in Figure 2, can be drawn for each of the contrasted groups, for example, high and low frontal. The effect of the conditions indicates that item-recognition learning is lowered in low-medial participants and that source-recognition learning is lowered in low-frontal participants. 2D-SDT does not require that the triangle of distribution means and distances be invariant for different populations. For 2D-SDT, the Glisky et al. results mean that brain damage affects learning on the two dimensions of the decision space in predictable ways.

Table 5
Forced-Choice Scores for Groups in Glisky et al. (1995)

Condition	Item	Source
High frontal	.88	.65
Low frontal	.83	.53
High medial	.91	.60
Low medial	.79	.59

In three experiments on generation effects, Jurica and Shimamura (1999) showed participants a series of questions and statements, accompanying each by one of three faces. The participants read the statements and answered the questions. Two successive memory tests were given, first a recall of the topics presented, in either the statements or questions, and then a multiple choice recognition test in which the studied questions and statements were presented with all three faces and the word *new* as alternatives. They found that on a recall test, the question topics were better recalled but on a source-recognition test, the statements were better identified than the questions. There are, however, several problems that should be considered with respect to the interaction. First, we do not have the same type of task used for both item memory and source memory. Item memory is tested by recall and source by recognition. The fact that the two methods can give different patterns of response is known. The most dramatic example of this difference is seen in word-frequency effects, in which low-frequency words are recognized better but recalled more poorly than high-frequency words. What is needed is a demonstration, using recognition tests, that generation and simple reading produce an interaction like that found by Jurica and Shimamura. Moreover, Hirshman and Bjork (1988) have discussed the complexity of generation operations and the variability of results they produce. This variability can be seen in Multhaup and Balota's (1997) healthy older participants. (These were compared with Alzheimer participants.) In part of the study, item and source recognition were compared. The generation condition (a cloze task) produced better item recognition (98%) than a reading task (87%). The generation task also produced better source recognition (86%) than the reading task (68%). There is no evidence of the crossover interaction of Jurica and Shimamura.

In another part of the experiment, the experimenter carried out the cloze task on some items as the participant listened. In the other, the participants did the cloze task. Here the participant-generated items were more accurately recognized (97%) than the experimenter-generated (93%) items, but the reverse held true for the source recognition (75% and 84%). Further empirical work is needed to clarify the effects in this area. As noted earlier in this article, if an interaction were established, it would be necessary to bring in the extension of 2D-SDT in Hilford et al. (2002). This extension, designed to explain the concavity of the z-ROC, consists of the addition of an attention component to the model.

A third study of interest here is one by Lindsay and Johnson (1991). In that study, deep and shallow encoding were used, and both source- and item-recognition accuracy were tested. Cards with words were presented to the left or right of the participants. One group (deep–deep) made up a sentence to go with each word on both sides as it was presented. The other group (deep–shallow) made up sentences for the words presented on the left (deep processing) and counted the number of *E*s in words presented on the right (shallow processing). The encoding operations were parallel to those used in Experiment 3 reported earlier in this article. The groups were divided into subgroups tested on item recognition and subgroups tested on source recognition. The accuracy scores (total percent correct) were as follows: For the deep–deep condition, source accuracy equals 0.65 and item accuracy equals 0.89; for the deep–shallow condition, source accuracy equals 0.83 and item accuracy equals 0.77. This crossover interaction can be covered, however, by 2D-SDT. If we take triangle III in Figure 2 and superimpose it on triangle II, matching Vertex A to A and N to N in each, we obtain distances that correspond to the probable distances in the Lindsay and Johnson data. (We say probable distances because Lindsay and Johnson use percent correct scores rather than d's.) The reason for the interaction is the differences between source learning in the deep–deep condition (II) and source learning in the deep–shallow condition (III). In the latter condition, the two sources have been made additionally distinct, in other words, separated on the source dimension.

It should be noted that in Experiments 1 and 4, source-recognition accuracy is greater than item-recognition accuracy. Such a relation is easily accommodated by 2D-SDT. It is, however, different from some findings in the literature. (See for example, the data of the three experiments in Slotnick et al., 2000.) One possible reason for the finding is in the procedure of Experiments 1 and 4. In both experiments, the participants first made an old–new discrimination and then for old responses, determined source. This procedure, called conditional source recognition, may introduce a bias that overestimates source-recognition accuracy (DeCarlo, 2003).

Alternative Approaches

There are two main alternative approaches to source recognition in the literature. Both have generated experimental work. One is the family of multinomial processing models of Batchelder and Riefer (1990). The other is the dual-process model of Yonelinas (1999). Neither approach can cover the regularities presented here and in Hilford et al. (2002), specifically the shape of the ROC. The disjunctions between those approaches and data are discussed fully in Hilford et al. They will be summarized briefly here.

The Batchelder and Riefer (1990) family of models is based on threshold mechanisms. Although these models have not been extended to cover confidence ratings, we can extend them and apply them to simplified source-recognition tasks such as that in Experiment 2. When we do that, we obtain the following rectilinear equation for the appropriate two high threshold model.

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

where A and B are two sources and d_1 is the probability of detecting the source of an item from A, d_2 is the probability of detecting the source of an item from B, and R_j is the cumulative distribution of ratings j. The corresponding equation's intercept in Equation 13 of Hilford et al. (2002) was incorrect, corrected here.

It has been demonstrated in all the relevant data in this article and in the experiments of Hilford et al. (2002), Qin et al. (2001), and Slotnick et al. (2000) that the source-recognition ROCs are convex as predicted by SDT and not rectilinear. The other alternative approach is the dual-process model of Yonelinas (1999). In this model, it is assumed that there are two different types of memory information used in source recognition. One type, familiarity, is processed by a signal detection mechanism. In isolation, it produces the standard convex ROC of SDT. The other type of information, recollection, is processed by a threshold mechanism. In isolation, recollection produces a rectilinear ROC.

The experiments reported here and in the earlier article (Hilford et al., 2002) were equal in familiarity. Analyses of the item-recognition d's in both articles supported the assertion that the two sources were indeed equal in familiarity. When the two sources are equal in familiarity, the dual-process model predicts a rectilinear ROC. See Yonelinas (1999).

In all three experiments of Hilford et al. (2002) and the five experiments reported in this article, the sources are of equal familiarity. Therefore, with dual-process theory, one would predict that they should have rectilinear ROCs. All eight experiments have convex ROCs that contradict the theory. Qin et al. (2001) and Slotnick et al. (2000) have made the same argument on the basis of their experiments.

The competence of SDT and the alternative threshold theories can be compared by examining, for each experiment, the variance for which they account. SDT predicts linear z-ROCs. Both alternative threshold theories predict linear, unconverted ROCs (not z-ROCs). The variances accounted for by SDT are listed in Appendix B in the next-to-last column. They are the R^2s obtained by a linear fit to the z-ROCs. The comparable variances accounted for by threshold models are listed next to them in the same appendix. They are the R^2s obtained by a linear fit to the ROCs. In nine of the twelve comparisons, SDT accounts for more of the variance.

Summary

In summary, the results presented here support the SDT explanation of source recognition. Most studies of source recognition have used some form of threshold theory based on Batchelder and Riefer's (1990) work (Dodson et al., 1998; Dodson & Shimamura, 2000; Ferguson, Hashtroudi, & Johnson, 1992; Johnson, Kounios, & Reeder, 1994; Meiser & Bröder, 2002). The present study offers an alternative that accounts for an average of 99% of the variance in the five experiments reported here. (See the linear R^2s of Appendix B.) The alternative has an advantage in that it brings item recognition and source recognition under a unified SDT structure. Item recognition has been extensively studied over the past 90 years and has been, as noted in the introduction to this article, fully incorporated in SDT models. The relations between the regularities of item recognition and source recognition recommend the use of a unified SDT structure. In particular, the demonstrated linkage between source and item recognition demand such a theory.

References

Allen, L. R., & Garton, R. F. (1968). The influence of word-knowledge on the word-frequency effect in recognition memory. *Psychonomic Science*, *10*, 401–402.

Anderson, J. R., & Milson, R. (1989). Human memory: An adaptive perspective. *Psychological Review*, *96*, 703–719.

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

Balota, D. A., & Neely, J. H. (1980). Test expectancy and word-frequency effects in recall and recognition. *Journal of Experimental Psychology*: *Human Learning and Memory*, *6*, 576–587.

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

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

Bornstein, B. H., & LeCompte, D. C. (1995). A comparison of item and source forgetting. *Psychonomic Bulletin & Review*, *2*, 254–259.

Bowles, N. L., & Glanzer, M. (1983). An analysis of interference in recognition memory. *Memory & Cognition*, *11*, 307–315.

Clark, E. O. (1980). Semantic and episodic memory impairment in normal and cognitively impaired elderly adults. In L. K. Obler & M. L. Albert (Eds.), *Language and communication in the elderly* (pp. 47–57). D. C. Health, Lexington, MA.

Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology*: *General, 104*, 268–294.

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

DeCarlo, L. T. (2003). Source monitoring and multivariate signal detection theory, with a model for selection. *Journal of Mathematical Psychology*, *47*, 292–303.

Dennis, S., & Humphreys, M. S. (2001). A context noise model of episodic word recognition. *Psychological Review*, *108*, 452–478.

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

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

Dorfman, D., Glanzer, M., & Kaufman, J. (1986). Aging effects on recognition memory when encoding and strategy are controlled. *Bulletin of the Psychonomic Society*, *24*, 172–174.

Duncan, M. (1999). The dimensionality of memory strength between levels of both serial position and word frequency category. *International Journal of Psychology*, *34*, 454–459.

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

Erber, J. T. (1974). Age differences in recognition memory. *Journal of Gerontology*, *29*, 177–181.

Erber, J. T. (1978). Age differences in a controlled-lag recognition memory task. *Experimental Aging Research*, *4*, 195–205.

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

Garton, R. F., & Allen, L. R. (1968). Familiarity and word recognition. *Quarterly Journal of Experimental Psychology*, *20*, 385–389.

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

Glanzer, M., & Adams, J. K. (1985). The mirror effect in recognition memory. *Memory & Cognition*, *13*, 8–20.

Glanzer, M., & Adams, J. K. (1990). The mirror effect in recognition memory: Data and theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*, 5–16.

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

Glanzer, M., & Bowles, N. (1976). Analysis of the word-frequency effect in recognition memory. *Journal of Experimental Psychology*: *Human Learning and Memory*, *2*, 21–31.

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.

Gordon, S. K., & Clark, W. C. (1974). Adult age differences in word and nonsense syllable recognition memory and response criterion. *Journal of Gerontology*, *29*, 659–665.

Gorman, A. M. (1961). Recognition memory for nouns as a function of abstractness and frequency. *Journal of Experimental Psychology*, *61*, 23–29.

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

Gregg, V. H. (1976). Word frequency, recognition and recall. In J. Brown (Ed.), *Recall and recognition* (pp. 183–216). London: Wiley.

Gronlund, S. D., & Elam, L. E. (1994). List-length effect: Recognition accuracy and variance of underlying distributions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 1355–1369.

Harkins, S. W., Chapman, C. R., & Eisdorfer, C. (1979). Memory loss and response bias in senescence. *Journal of Gerontology*, *34*, 66–72.

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

Heathcote, A. (2003). Item recognition memory and the receiver operating characteristic. *Journal of Experimental Psychology*: *Learning, Memory, and Cognition*, *29*, 1210–1230.

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.

Hilford, A., Glanzer, M., & Kim, K. (1997). Encoding, repetition, and the mirror effect in recognition memory: Symmetry in motion. *Memory & Cognition*, *25*, 593–605.

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

Hirshman, E., & Bjork, R. A. (1988). The generation effect: Support for a two-factor theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*, 484–494.

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., & 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.

Jurica, P. J., & Shimamura, A. P. (1999). Monitoring item and source information: Evidence for a negative generation effect in source memory. *Memory & Cognition*, *27*, 648–656.

Kausler, D. H., & Kleim, D. M. (1978). Age differences in processing relevant versus irrelevant stimuli in multiple-item recognition learning. *Journal of Gerontology*, *33*, 87–93.

Kim, K., & Glanzer, M. (1993). Speed versus accuracy instructions, study time, and the mirror effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*, 638–652.

Kinsbourne, M., & George, J. (1974). The mechanism of the word-frequency effect on recognition memory. *Journal of Verbal Learning and Verbal Behavior*, *13*, 63–69.

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

Lawrence, D. M., & Banks, W. P. (1973). Accuracy of recognition memory for common sounds. *Bulletin of the Psychonomic Society*, *1*, 298–300.

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

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

McCormack, P. D., & Swenson, A. L (1972). Recognition memory for common and rare words. *Journal of Experimental Psychology*, *95*, 72–77.

Meiser, T., & Bröder, A. (2002). Memory for multidimensional source information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 116–137.

Multhaup, K. S., & Balota, D. A. (1997). Generation effects and source memory in healthy older adults and in adults with dementia of the Alzheimer type. *Neuropsychology*, *11*, 382–391.

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 Monographs*, *76*, (*1*, Pt. 2).

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.

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., & Murdock, B. B., Jr. (1976). Retrieval processes in recognition memory. *Psychological Review*, *83*, 190–214.

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

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.

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

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 a normal subject. *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.

Spencer, W. D., & Raz, N. (1995). Differential effects of aging on memory for content and context: A meta-analysis. *Psychology and Aging*, *10*, 527–539.

Strong, E. K., Jr. (1912). The effect of length of series upon recognition memory. *Psychological Review*, *19*, 447–462.

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

Thomas, J. P. (1985). Detection and identification: how are they related? *Journal of the Optical Society of America, 2*, 1457–1467.

Troyer, A. K., Winocur, G., Craik, F. I. M., & Moscovitch, M. (1999). Source memory and divided attention: Reciprocal costs to primary and secondary tasks. *Neuropsychology*, *13*, 467–474.

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

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

(*Appendixes follow*)

Appendix A

Group Data: Number of Responses in Each Confidence Rating Summed Across Participants

	Confidence ratings					
Variable	1	2	3	4	5	6
1. Repetition						
Single						
Male voice	393	258	237	184	145	103
Female voice	82	164	217	206	259	405
Repetition						
Male voice	759	307	254	139	187	102
Female voice	122	154	192	216	341	698
Experiment 2. List length						
Long						
Top	213	175	247	304	142	39
Bottom	39	124	232	302	196	227
Short						
Top	336	173	216	220	120	55
Bottom	53	130	198	246	186	307
Experiment 3. Encoding						
Letter						
Male voice	154	114	179	132	84	38
Female voice	44	92	149	169	120	125
Frequency						
Male voice	156	136	138	134	102	33
Female voice	23	90	127	156	136	169
Experiment 4. Word frequency						
High						
Top	186	149	164	107	95	41
Bottom	38	122	130	145	156	166
Low						
Top	307	227	183	166	130	54
Bottom	57	112	174	163	222	314
Experiment 5. Aging (source)						
Old						
Male	124	104	127	95	83	43
Female	30	74	84	118	132	138
Young						
Male	411	138	183	150	86	40
Female	50	85	150	167	165	391
Experiment 5. Aging (item)						
Old						
Studied	461	162	100	128	192	108
New	58	80	111	167	340	397
Young						
Studied	1093	236	179	228	205	75
New	89	114	222	471	625	495

Appendix B

Group ROC and z-ROC Statistics

Experiment and variable	ROC	z-ROC			SDT linear z-ROC R^2	Threshold linear ROC R^2
	Quadratic constant	Linear slope	Linear intercept	Quadratic constant		
1. Repetition						
Single	−0.97	0.95	0.88	0.10	0.995	0.907
Repeat	−1.66	1.00	1.29	0.07	0.997	0.919
2. List length						
Long	−0.42	1.00	0.76	0.23	0.969	0.993
Short	−0.60	0.95	0.92	0.24	0.972	0.987
3. Encoding						
Letter	−0.55	0.99	0.63	0.13	0.990	0.985
Frequency	−0.59	0.95	0.81	0.21	0.971	0.986
4. Word frequency						
High	−0.71	0.94	0.77	0.13	0.986	0.974
Low	−1.07	1.02	1.02	0.14	0.990	0.956
5. Aging (source)						
Old	−0.92	0.97	0.72	0.07	0.995	0.979
Young	−1.17	1.02	1.33	0.29	0.976	0.966
5. Aging (item)						
Old	−0.95	0.76	0.97	0.08	0.994	0.947
Young	−0.75	0.69	1.26	0.07	0.993	0.914

Note. ROC = receiver operating characteristic; z-ROC = receiver operating characteristic plotted in standard scores; SDT = signal detection theory.

Received April 22, 2003
Revision received June 8, 2004
Accepted June 11, 2004 ■