Journal of Experimental Psychology:
Learning, Memory, and Cognition
1996, Vol. 22, No. 1, 197–215

Copyright 1996 by the American Psychological Association, Inc.
0278-7393/96/$3.00

Source Discrimination, Item Detection, and Multinomial Models of Source Monitoring

Ute J. Bayen
The University of Memphis

Kevin Murnane
The Pennsylvania State University

Edgar Erdfelder
University of Bonn

Source monitoring refers to the discrimination of the origin of information. Multinomial models of source monitoring (W. H. Batchelder & D. M. Riefer, 1990) are theories of the decision processes involved in source monitoring that provide separate parameters for source discrimination, item detection, and response biases. Three multinomial models of source monitoring based on different models of decision in a simple detection paradigm (one-high-threshold, low-threshold, and two-high-threshold models) were subjected to empirical tests. With a 3 (distractor similarity) × 3 (source similarity) factorial design, the effect of difficulty of item detection and source discrimination on corresponding model parameters was examined. Only the source-monitoring model that is based on a two-high-threshold model of item recognition provides an accurate analysis of the data. Consequences for the use of multinomial models in the study of source monitoring are discussed.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Source monitoring refers to the discrimination of the origin of information. In the typical source-monitoring task, items of information are presented from two or more sources and the correct mapping between source and item of information must be remembered at a later time. Trying to remember which journal this article appeared in several months from now is an example. Over the last 15 years, experimental paradigms using source-monitoring tasks have gained increasing popularity in many fields of psychology including basic and applied memory research, psycholinguistics, social psychology, developmental psychology, educational psychology, and neuropsychology. Source-monitoring paradigms have been used to address questions regarding eyewitness testimony (Lindsay, 1990; Lindsay & Johnson, 1989; Zaragoza & Lane, 1994), cryptomnesia (Brown & Halliday, 1991; Marsh & Bower, 1993), bilinguality (for a review, see Gerard & Scarborough, 1989), and the relationship between the credibility of information and its source (Begg, Anas, & Farinacci, 1992). Developmental psychologists and gerontologists have compared source-monitoring performance in different age groups (e.g., Ferguson, Hashtroudi, & Johnson, 1992; Gopnik & Graf, 1988; Hashtroudi, Johnson, Vnek, & Ferguson, 1994; Lindsay, Johnson, & Kwon, 1991; Spencer & Raz, 1994); whereas neuropsychologists have been interested in source monitoring in different clinical populations (Janowsky, Shimamura, & Squire, 1989; Mitchell, Hunt, & Schmitt, 1986; Taylor, Saint-Cyr, & Lang, 1990). For a comprehensive review of source-monitoring research in different fields, see Johnson, Hashtroudi, and Lindsay (1993).

Although source-monitoring studies have been of interest for over a decade, extended theories of source monitoring have been developed only recently. Johnson et al. (1993) presented a general theoretical framework for the memory processes involved in source monitoring. Batchelder and his colleagues (Batchelder & Riefer, 1990; Batchelder, Hu, & Riefer, 1994; Batchelder, Riefer, & Hu, 1994; Riefer, Hu, & Batchelder, 1994) have pioneered the development of multinomial models of the decision processes involved in source monitoring. There are many well-known benefits to be derived from explaining phenomena with carefully formulated formal theories (see Hintzman, 1991, for a recent discussion). One of the advantages of multinomial theories of source monitoring is that they provide independent, theoretically motivated parameters for the measurement of different factors that can contribute to performance in a source-monitoring task. In experimental source-monitoring tasks, it is common for participants to be required to both detect the difference between old and new items and discriminate between different sources for old items. Use of this type of task raises important measurement issues because commonly used empirical measures of source monitoring confound memory for the item with memory for the source. That is, commonly used measures of memory for the source are sensitive to experimental manipulations that affect memory for

Ute J. Bayen, Department of Psychology, The University of Memphis; Kevin Murnane, Department of Psychology, The Pennsylvania State University; Edgar Erdfelder, Department of Psychology, University of Bonn, Bonn, Germany.

The research reported in this article was supported by National Science Foundation Grant SBR-9319549, a fellowship from the German National Fellowship Foundation, an Alumni Faculty Research Fellowship and a Research Initiation Grant from The Pennsylvania State University, and a Center of Excellence Grant awarded to the Department of Psychology, The University of Memphis by the state of Tennessee. We thank William Batchelder, Douglas Hintzman, and Steve Lindsay for helpful comments on a version of the article.

Correspondence concerning this article should be addressed to Ute J. Bayen, Department of Psychology, The University of Memphis, Memphis, Tennessee 38152 or to Kevin Murnane, Department of Psychology, 614 Moore Building, The Pennsylvania State University, University Park, Pennsylvania 16802. Electronic mail may be sent via Internet to Ute J. Bayen at ubayen@irvin.psyc.memphis.edu or to Kevin Murnane at kxm20@psuvm.psu.edu.

the item but do not affect memory for the source. In addition, most commonly used empirical measures cannot separate source sensitivity from response bias with regard to the source (Batchelder & Riefer, 1990; Batchelder, Hu, & Riefer, 1994; Murnane & Bayen, in press; Riefer et al., 1994). Multinomial models of source monitoring solve this problem by providing independent parameters for the measurement of item memory, source memory, and various types of response bias. In this article, we compare the multinomial models presented by Batchelder and Riefer (1990) and Batchelder, Riefer, and Hu (1994), identify failings of both models, and present a multinomial model of source monitoring that avoids these problems.[1]

Because performance in a source-monitoring task involves both memory for the item and memory for the source, tests of source-monitoring theories must evaluate the ability of the theory to account for both types of memory. Fitting a model to data derived from published source-monitoring studies is not likely to provide an adequate test of the model because most source-monitoring studies have only manipulated factors expected to affect source discrimination; factors affecting memory for the item are usually not examined. One of our goals in the present article is to provide a data set that can serve as a basis for rigorous tests of source-monitoring theories. To this end, we present an experimental paradigm in which factors affecting memory for the item and memory for the source are independently manipulated and shown to have independent effects on performance. The data from this paradigm are then used to test the three multinomial models of source monitoring mentioned above.

Studies of source monitoring have used different kinds of experimental tasks. In the most frequently used experimental paradigm, a number of items (such as words, sentences, etc.) are presented to participants by two different sources, A and B. Sources have been defined as different people, presentation media, study lists, sensory modalities, and so forth. After items have been presented during the learning phase of the experiment, participants receive a three-alternative source-monitoring test in which each test item must be identified as having been originally presented by Source A, as having been originally presented by Source B, or as a new item that was not presented during learning.

The data from a typical source-monitoring experiment of this type can be summarized with a 3 (source of item) × 3 (participant's response) frequency matrix as shown in Table 1, in which Y_{ij} is the frequency of response j to items of type i (adapted from Batchelder & Riefer, 1990). Responses in a typical source-monitoring experiment may be influenced by the participant's ability to recognize an item as old or new, the participant's ability to discriminate sources, and by various forms of response bias. Analysis of source-monitoring data therefore poses the challenge of disentangling the contributions of these different influences.

Table 1
Data From a Typical Source-Monitoring Experiment

	Response		
Source	"A"	"B"	"N"
A	Y_{AA}	Y_{AB}	Y_{AN}
B	Y_{BA}	Y_{BB}	Y_{BN}
N	Y_{NA}	Y_{NB}	Y_{NN}

Note. A = Source A; B = Source B; N = distractor item; Y_{ij} = the frequency of responses of type j to items of type i.

How best to measure performance in a source-monitoring task has engendered considerable controversy (Batchelder & Riefer, 1990; Batchelder, Riefer, & Hu, 1994; Kinchla, 1994; Murnane & Bayen, in press; Riefer et al., 1994). In simple detection paradigms that make use of a single-dimensioned decision space, discrimination of old and new items independent of biases to respond "old" or "new" is usually measured by using the sensitivity measure derived from signal detection theory (SDT), d'. However, Batchelder, Riefer, and Hu (1994) pointed out that one of the assumptions involved in the typical calculation of d', that the variances of the target and distractor distributions are equal, may not be met in most simple recognition paradigms (Gronlund & Elam, 1994; Ratcliff, Sheu, & Gronlund, 1992). Batchelder, Riefer, and Hu (1994) also questioned whether d' is an appropriate measure for detection and recognition paradigms involving multidimensional stimuli. In addition, Thomas and Olzak (1992) showed that d' can be a biased measure of performance if it is applied to tasks requiring simultaneous detection and discrimination such as the source-monitoring task described above. Despite these critiques, we use d' as a measure of item detection for two reasons. First, as discussed below, the pattern of data found in our experiment is difficult to reconcile with a multidimensional model as discussed by Thomas and Olzak. Second, comparison of the frequently used d' with recognition measures derived from multinomial models of source monitoring is of some interest, even if d' should turn out to be a biased measure of item detection in source-monitoring experiments. We calculated d' from the response frequencies shown in Table 1 on the basis of the hit rate (HR),

$$\mathrm{HR} = \frac{Y_{AA} + Y_{AB} + Y_{BA} + Y_{BB}}{Y_{AA} + Y_{AB} + Y_{AN} + Y_{BA} + Y_{BB} + Y_{BN}}, \quad (1)$$

and the false-alarm rate (FAR),

$$\mathrm{FAR} = \frac{Y_{NA} + Y_{NB}}{Y_{NA} + Y_{NB} + Y_{NN}}. \quad (2)$$

Measurement of source discrimination in a source-monitoring task is also controversial. Murnane and Bayen (in press) have shown that most of the commonly used measures of source monitoring confound item recognition with source discrimination in most circumstances. They also demonstrated that the average conditional source identification measure (ACSIM) provides a measure of source discrimination that is independent of item recognition in many circumstances in

[1] During the review process we learned that the two-high-threshold model of source monitoring that we present in detail in this article has also been considered by Batchelder, Hu, and Riefer (1994, cf. Equation 1.13). Our formulation, development, and theoretical and empirical analysis of this model was conducted before our learning of Batchelder, Hu, and Riefer's work.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

which other commonly used measures are not. However, ACSIM is not independent of item recognition in some circumstances when targets are identified as old on the basis of guessing. Because ACSIM is not a commonly used measure, we present the calculational formula for a two-source experiment expressed in terms of the response frequencies shown in Table 1,

$$\mathrm{ACSIM} = \frac{\dfrac{Y_{\mathrm{AA}}}{Y_{\mathrm{AA}} + Y_{\mathrm{AB}}} + \dfrac{Y_{\mathrm{BB}}}{Y_{\mathrm{BB}} + Y_{\mathrm{BA}}}}{2}. \quad (3)$$

We use ACSIM to measure source discrimination because, compared with other empirical measures of source monitoring, it is independent of item recognition in the widest range of circumstances.[2]

Three Multinomial Models of Source Monitoring

Multinomial models are based on the assumption that the probabilities of different behavioral responses can be mapped onto the probabilities of different underlying cognitive states. The probability that the participant is in one of these cognitive states is represented in the model with a hypothetical parameter that can be estimated from data by using well-understood statistical methods. Riefer and Batchelder (1988) argued that multinomial models are superior to general statistical models such as the general linear model of analysis of variance (ANOVA) because the multinomial models rest on specific assumptions about the nature of cognitive states and thus enable measurement of the effects of cognitive processes on behavior. They also argued that, in some circumstances, multinomial models may be preferable to strong theoretical models such as SAM (the search of associative memory model; Raaijmakers & Shiffrin, 1981) or Minerva 2 (Hintzman, 1988) because the mathematics of multinomial models are relatively simple and support mathematical derivations of parameter estimates rather than the generation of parameter estimates through computer simulations.[3]

Batchelder and his colleagues have developed several multinomial models for a variety of source-monitoring task domains. Batchelder and Riefer (1990) and Batchelder, Riefer, and Hu (1994) presented different multinomial models that apply to cases in which participants must discriminate between two sources. Riefer et al. (1994) have developed a multinomial model for analyzing data from experiments with three sources, and Batchelder, Hu, and Riefer (1994) presented a model that can be applied with an arbitrary number of sources. In this article we are concerned with multinomial models that apply to the discrimination of two sources.

Multinomial models of source monitoring are variants of the general class of threshold theories that have been developed for simple detection and recognition paradigms. Threshold theories (Krantz, 1969; Luce, 1963a, 1963b) are based on the assumption that the decision space in a detection paradigm can be divided into an arbitrary number of discrete states. Threshold theories differ over the number of states that define the decision space and over the probabilities of these states given a specific type of stimulus (e.g., signal, noise, target, distractor, and so forth). Multinomial theories of source monitoring define two decision spaces: one that describes the item detection component of the task, and one that describes the source discrimination component. The three multinomial theories we consider in this article are all based on a two-high-threshold (2HT) model of the source discrimination component of the source-monitoring task. In a 2HT model, there are two thresholds that divide the decision space into three discrete areas. In a 2HT model of source discrimination, if one threshold is crossed, the source is identified as Source A, if the other threshold is crossed, the source is identified as Source B, and if neither threshold is crossed, the source is identified as Source A with a guessing probability and is identified as Source B with the complementary guessing probability. It is assumed that only items that came from Source A can cross the Source A threshold, and only items that came from Source B can cross the Source B threshold. While all models discussed in this article are based on a 2HT model of source discrimination, they differ in how they model the decision space for item detection.

A One-High-Threshold Multinomial Model of Source Monitoring

Batchelder and Riefer (1990) presented a multinomial model of source monitoring that includes a one high-threshold (1HT) model of item detection. In 1HT models there is a single high threshold that divides the decision space into two discrete areas. If the threshold is crossed on presentation of a test item, the item is detected as old; if the threshold is not crossed, the item is said to be in an undetected state, and the participant responds either "old" or "new" on the basis of a guessing probability. In 1HT models it is assumed that only old items can cross the high threshold. All new items and the old items that do not cross the threshold are categorized as *old* or *new* on the basis of guessing.

To maintain the distinction between a 1HT model of simple item detection and Batchelder and Riefer's (1990) model of source monitoring, we refer to the Batchelder and Riefer (1990) model as a 1HT model of source monitoring or 1HTSM. This model is illustrated in Figure 1, and the parameters of the model are given in Table 2.

Predictions are derived from the model as follows. Assume an item that originated from Source A is presented at test and refer to the first decision tree in Figure 1. The participant will recognize the test item as a target with probability D_1. With probability d_1, the recognized item will correctly be identified as having originated from Source A; with the complementary

[2] A detailed discussion of the strengths and weaknesses of different empirical measures of source discrimination under different sets of assumptions about the memory and decision processes involved in the performance of a source-monitoring task is beyond the scope of this article (however, see Murnane & Bayen, in press). Data analyses using alternative empirical measures produced patterns of results that were identical to the analyses using ACSIM that are reported in this article.

[3] We note that mathematical derivation of predictions is possible in some circumstances for the strong theoretical models as well. For example, closed-form solutions can be derived for recognition within the SAM model (see, for example, Murnane & Shiffrin, 1991).

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

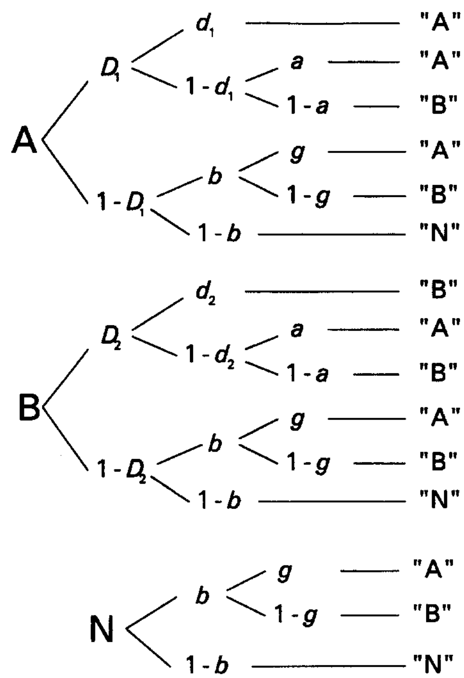

Figure 1. One-high-threshold multinomial model of source monitoring. A = Source A item; B = Source B item; N = distractor item; D_1 = probability of detecting an item from Source A; D_2 = probability of detecting an item from Source B; d_1 = probability of correctly discriminating the source of an item from Source A; d_2 = probability of correctly discriminating the source of an item from Source B; a = probability of guessing that a detected item is from Source A; b = probability of guessing an item is old; g = probability of guessing that an undetected item is from Source A. From "Multinomial Processing Models of Source Monitoring," by W. H. Batchelder and D. M. Riefer, 1990, *Psychological Review, 97,* p. 551. Copyright 1990 by the American Psychological Association. Adapted with permission of the authors.

probability $1 - d_1$, the participant will not be able to identify the correctly recognized item as having originated from Source A and will, therefore, guess the source of the item. With probability a, the participant will guess that the item originated from Source A; with the complementary probability $1 - a$, the participant will guess that the item originated from Source B. If the test item is not recognized, with probability $1 - D_1$, the participant will, with probability b, guess that it is a target item. Given this decision, the participant will then guess, with probability g, that the item originated from Source A, or will guess, with probability $1 - g$, that the item originated from Source B. If the participant does not guess that the unrecognized item is a target, the item will, with probability $1 - b$, be classified as a distractor. The probabilities for each of the possible responses for items from Source A are arrived at by summing the probabilities from the branches of the tree diagram that lead to the appropriate response. These probabilities are given by Equations 4, 5, and 6:

$$P(\text{“A”}\,|\,\text{A}) = D_1 d_1 + D_1(1 - d_1)a + (1 - D_1)bg, \qquad (4)$$

$$P(\text{“B”}\,|\,\text{A}) = D_1(1 - d_1)(1 - a) + (1 - D_1)b(1 - g), \qquad (5)$$

and

$$P(\text{“N”}\,|\,\text{A}) = (1 - D_1)(1 - b), \qquad (6)$$

where the letter in quotation marks indicates the participant's response that the item originated from Source A ("A"), Source B ("B"), or is a new item ("N"). Model equations for items that originated from Source B and for distractor items are arrived at in a similar manner by using the appropriate decision trees in Figure 1.

The model, as depicted in Figure 1, is not technically identifiable because the model has seven free parameters, and there are only six independent model equations for the typical source-monitoring task. Thus, the parameters of the full seven-parameter model cannot be estimated from data. However, as discussed in detail by Batchelder and Riefer (1990), it is possible to impose equality restrictions on the parameters that define globally identifiable submodels. For these submodels, well-known statistical techniques of parameter estimation (e.g., the maximum-likelihood method) may be used to derive estimates of parameters on the basis of the response frequencies shown in Table 1 (see also Hu & Batchelder, 1994).

It has been pointed out many times that 1HT theories for simple detection and recognition paradigms make predictions about the shape of the receiver-operating characteristic (ROC) that are contrary to data (e.g., Green & Swets, 1966; Kintsch, 1970; Luce, 1963a; Murdock, 1974). Kinchla (1994) has argued that because the Batchelder and Riefer (1990) model includes a 1HT model of item detection, it is inadequate as a theory of source monitoring. In response to Kinchla's criticism, Batchelder, Riefer, and Hu (1994) stated that the inadequacy of the 1HT model as a model for simple detection and recognition paradigms does not imply the inadequacy of the high-threshold assumption in multinomial models of source monitoring. They also presented a low-threshold multinomial model of source monitoring to which we now turn.

A Low-Threshold Multinomial Model of Source Monitoring

In low-threshold (LT) models, there is a single low threshold that divides the decision space into two discrete areas and it is assumed that both old and new items can cross the low

Table 2
Parameters for the Batchelder and Riefer (1990) 1HTSM Model

Parameter	Meaning
D_1	Probability of detecting an item from Source A
D_2	Probability of detecting an item from Source B
d_1	Probability of correctly discriminating the source of an item from Source A
d_2	Probability of correctly discriminating the source of an item from Source B
b	Probability of guessing an item is old
a	Probability of guessing a detected item is from Source A
g	Probability of guessing an undetected item is from Source A

Note. 1HTSM = one-high-threshold source-monitoring model.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

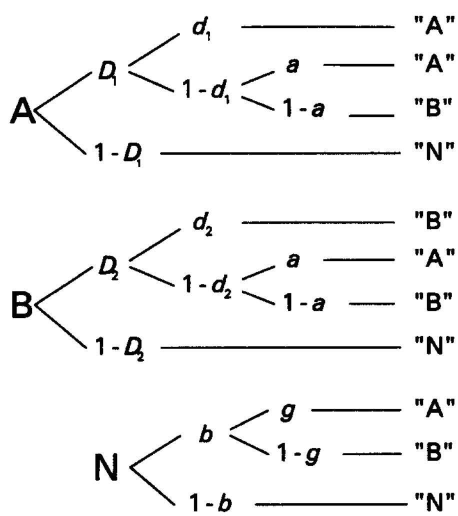

Figure 2. Low-threshold multinomial model of source monitoring (Batchelder, Riefer, & Hu, 1994). A = Source A item; B = Source B item; N = distractor item; D_1 = probability of detecting an item from Source A; D_2 = probability of detecting an item from Source B; d_1 = probability of correctly discriminating the source of an item from Source A; d_2 = probability of correctly discriminating the source of an item from Source B; a = probability of guessing that a detected item is from Source A; b = probability of a false alarm; g = probability of guessing that a distractor item is from Source A.

threshold. There are two common variants of LT models (Macmillan & Creelman, 1991). In models that use an "upper limb" strategy, items that cross the threshold are detected as old, and items that fail to cross the threshold are in an undetected state and are identified as old or new with a guessing probability. In models that use a "lower limb" strategy, items that cross the threshold are identified as new, and items that fail to cross the threshold are undetected and are responded to with a guess.

Batchelder, Riefer, and Hu (1994) presented a low-threshold model of source monitoring (LTSM) that includes a variant of the typical LT models described above. There is no guessing state for item detection in the Batchelder, Riefer, and Hu (1994) model; items that cross the threshold are identified as old, and items that fail to cross the threshold are identified as new. The model is shown in Figure 2. Parameters have the same interpretation as in the 1HTSM model with the exception that in the LTSM model, the b parameter indicates the probability that a new item is detected as old, that is, the probability of a false alarm. Predictions are derived in the same manner as illustrated for the 1HTSM model. Like the 1HTSM model, the LTSM model, as depicted in Figure 2, is not identifiable. Again, however, identifiable submodels may be derived by imposing equality restrictions on parameters. Parameters are estimated by using the same techniques as for the 1HTSM model.

Low-threshold models for simple detection and recognition paradigms that combine an upper and a lower limb strategy are able to predict ROC data better than their 1HT counterparts (Green & Birdsall, 1978; Luce, 1963b). Note, however, that the LTSM variant suggested by Batchelder, Riefer, and Hu (1994) is not as flexible as Luce's (1963b) model because it does not include upper and lower limb strategies. Moreover, the additional constraints included in the Batchelder, Riefer, and Hu (1994) LTSM model render the model less attractive as a general model of source monitoring. In the 1HTSM model, the b parameter provides an independent measure of recognition bias. In the LTSM model, the b parameter measures the probability that new items are incorrectly identified as old, and the model does not account for possible response biases or guessing processes with respect to Source A and Source B items. This can easily be seen by comparing Figures 1 and 2. In the 1HTSM model, when a target is undetected (with probability $1 - D_1$), the participant guesses old with probability b and new with probability $1 - b$. Under the same circumstances in the LTSM model, the participant identifies the test item as new with probability 1.0. The lack of a recognition bias parameter weakens the LTSM model as a general purpose model of source monitoring.

A Two-High-Threshold Multinomial Model of Source Monitoring

Like LT models, 2HT models for simple detection and recognition paradigms do not fare as poorly as 1HT models when compared against ROC data (Macmillan & Creelman, 1990; Macmillan & Kaplan, 1985; Snodgrass & Corwin, 1988; Swets, 1986). Two-high-threshold models of the item detection component of the source-monitoring task are similar to the 2HT model of the source discrimination component described earlier. There are two thresholds that divide the decision space into three discrete areas that correspond to detect as old, detect as new, and undetected. If either threshold is crossed on presentation of a test item, the item is detected as either old or new, depending on which threshold was crossed; if neither threshold is crossed, the item is in an undetected state and the participant responds either "old" or "new," depending on a guessing probability. In 2HT models, it is assumed that only old items can cross the detect-as-old threshold and only new items can cross the detect-as-new threshold.

A 2HT source-monitoring model (2HTSM) can be constructed by adding a parameter that indicates the probability that a new item will be detected as new to the 1HTSM model of Batchelder and Riefer (1990). We label this added parameter D_3. The 2HTSM model is shown in Figure 3. Predictions are derived in the same manner as for the 1HTSM and LTSM models.

The 1HTSM model illustrated in Figure 1 can be derived from the 2HTSM model shown in Figure 3 by imposing the restriction that $D_3 = 0$. A 1HT model may always be regarded as a special case of a 2HT model in which the probability of crossing the second high threshold is zero for all classes of items. Thus, the submodels of the 1HTSM model illustrated in Figure 2 of Batchelder and Riefer (1990, p. 552) may also be regarded as submodels of the 2HTSM model. In contrast to the 1HTSM and 2HTSM models, the LTSM and 2HTSM models are unrelated in the sense that there are no sets of parameter fixations or equality restraints that transform one of the models into the other.

Figure 4 (analogous to Figure 2 in Batchelder & Riefer, 1990, p. 552) shows a nested hierarchy of all identifiable

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

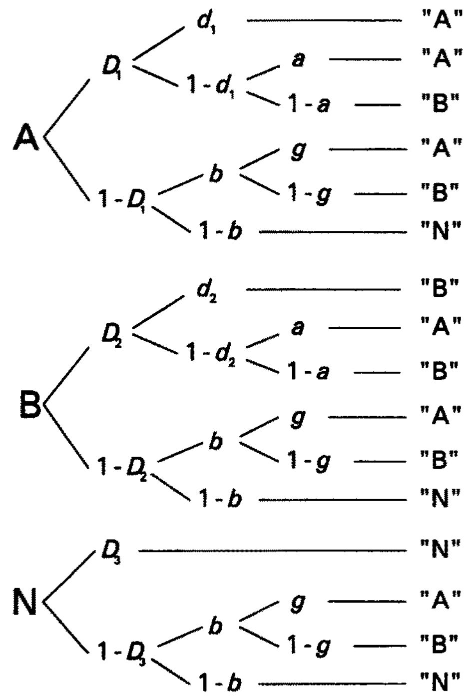

Figure 3. Two-high-threshold multinomial model of source monitoring. A = Source A item; B = Source B item; N = distractor item; D_1 = probability of detecting an item from Source A; D_2 = probability of detecting an item from Source B; D_3 = probability of detecting that a distractor is new; d_1 = probability of correctly discriminating the source of an item from Source A; d_2 = probability of correctly discriminating the source of an item from Source B; a = probability of guessing that a detected item is from Source A; b = probability of guessing an item is old; g = probability of guessing that an undetected item is from Source A.

submodels that may be obtained by means of equality restrictions on the parameters of the 2HTSM model. The statistical analysis of these submodels proceeds in the same way as previously discussed for the 1HTSM and LTSM models.

The fact that the 1HTSM, LTSM, and 2HTSM models, as presented in Figures 1, 2, and 3, respectively, contain different numbers of parameters does not prevent an empirical comparison of these models. When fitting the models to data, we show how all three models can be reduced to versions with equal numbers of parameters that can be compared directly.

Overview of the Experiment

In addition to testing comparable, identifiable submodels of the three multinomial models described above, our goal in designing this experiment was to develop an empirical paradigm that would provide a rich source of data for testing any theory of source monitoring. One criterion that an acceptable theory of source monitoring must meet is the ability to account for both memory for the item and memory for the source. For this reason, variables affecting item recognition (semantic similarity of targets and distractors) and source discrimination (perceptual similarity of sources) were directly and independently manipulated in the experiment reported below. To produce a rich data set that would provide a rigorous test of theory, we manipulated each variable across three levels to produce a completely crossed 3 × 3 design. For a theory to perform adequately, it must accurately predict the effects of the experimental manipulations across all nine cells of the design.

Prior empirical tests of the 1HTSM (Batchelder & Riefer, 1990) and LTSM (Batchelder, Riefer, & Hu, 1994) models have involved use of the models to reanalyze data sets from previously published source-monitoring experiments. Goodness-of-fit measures indicated good fits for the analyzed data sets. The results of parameter estimates were plausible, and analyses with both models led to very similar conclusions. However, because source memory experiments typically involve manipulations designed to affect source discrimination and not item recognition, the ability of the models to measure correctly both item detection and source discrimination has not adequately been tested. The following experiment provided a test of the ability of the models to measure both memory for the item and memory for the source.

Method

Participants

Two hundred and sixteen volunteers were recruited from introductory psychology courses at The Pennsylvania State University. They

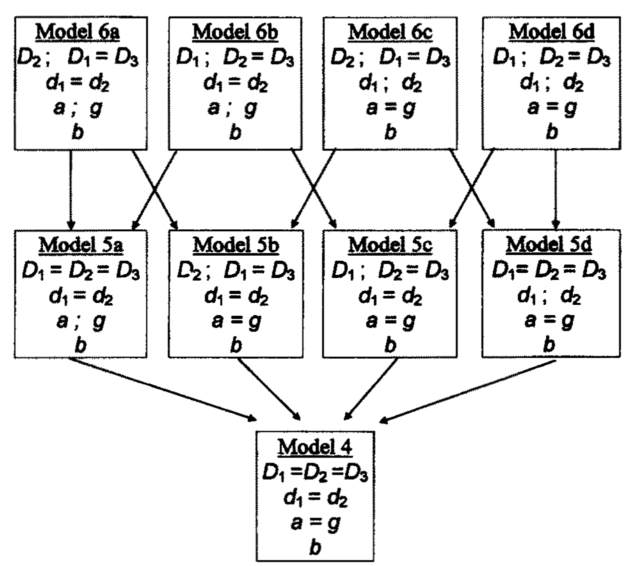

Figure 4. A nested hierarchy of all identifiable submodels of the two-high-threshold source-monitoring model. Arrows indicate proper subset relations between models which correspond to equality restrictions between parameters. An equal sign indicates that two parameters are constrained to be equal, and a semicolon separates parameters that are allowed to vary in a given model. D_1 = probability of detecting an item from Source A; D_2 = probability of detecting an item from Source B; D_3 = probability of detecting that a distractor is new; d_1 = probability of correctly discriminating the source of an item from Source A; d_2 = probability of correctly discriminating the source of an item from Source B; a = probability of guessing that a detected item is from Source A; g = probability of guessing that an undetected item is from Source A; b = probability of guessing an item is old.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

received class credit as compensation for their participation in the study. All participants were native speakers of English. Participants were randomly assigned to the nine experimental cells so that each cell contained 24 participants. Six additional participants were tested to replace the data from participants who did not complete the experiment because of either technical failure or failure to follow instructions.

Design

Participants were asked to place themselves in the role of a detective trying to locate a missing person named Mary. Fifty-two items of information about Mary were presented in the context of coherent narrative stories told by two different sources. The narratives are reproduced in Appendix A. Each source was identified as an individual who has a personal relationship with Mary such as her husband or brother. Sources differed in name, voice, picture, and, in the low source-similarity condition, sex.

Source discrimination was affected by manipulating the perceptual similarity of the two sources (Ferguson et al., 1992; Lindsay et al., 1991). Item recognition was affected by manipulating the semantic similarity of targets and distractors on a three-alternative source-monitoring test. Each of the independent variables was manipulated across three levels to produce a fully crossed 3 (distractor similarity) × 3 (source similarity) between-subjects factorial design. Accuracy of item detection was predicted to decline as distractor similarity increased; accuracy of source discrimination was predicted to decline as source similarity increased. In accordance with the assumption made by all of the multinomial models that independent cognitive processes underlie item recognition and source discrimination, it was expected that the manipulation of source similarity would affect source discrimination but not item recognition. Likewise, the manipulation of distractor similarity was expected to affect item recognition but not source discrimination. The distractor similarity manipulation was expected to affect item recognition by increasing FAR; HR was expected to remain unchanged.

Materials

Four black-and-white pictures representing the sources used in the experiment were constructed by using the "Mac a Mug" computer program. High-similarity faces shared more facial features than low-similarity faces. The four faces used in the experiment are shown in Figure 5. All participants received information from Jack, who was identified as Mary's brother. Participants in high source-similarity conditions also received information from John, who was identified as Jack's twin brother. The second source in the medium source-similarity conditions was Michael, Mary's husband. Susan, a friend of Mary's, was the second source in the low-similarity conditions.

Jack's narrative was presented in one male voice, John or Michael's narrative was presented in a different male voice, and Susan's narrative was presented in a female voice. The two male voices were sufficiently different to be easily distinguished.

Pilot studies in which participants were asked to assess the relative similarity of a target word to each of three related words were used to construct sets of four words composed of a target item (e.g., "lemons") coupled with high- ("limes"), medium- ("oranges"), and low- ("bananas") similarity distractors. The 52 item quadruplets used in the experiment are reproduced in Appendix B. The item quadruplets were randomly divided into two groups corresponding to the two narratives told by Sources A and B. Each target item was embedded in a sentence (e.g., "Mary hates lemons") in such a way that the target could be replaced with any one of its paired distractors with little, if any, modification of basic sentence structure. The resulting 52 sentences (two sets of 26 sentences) were used to construct the two narratives about Mary.

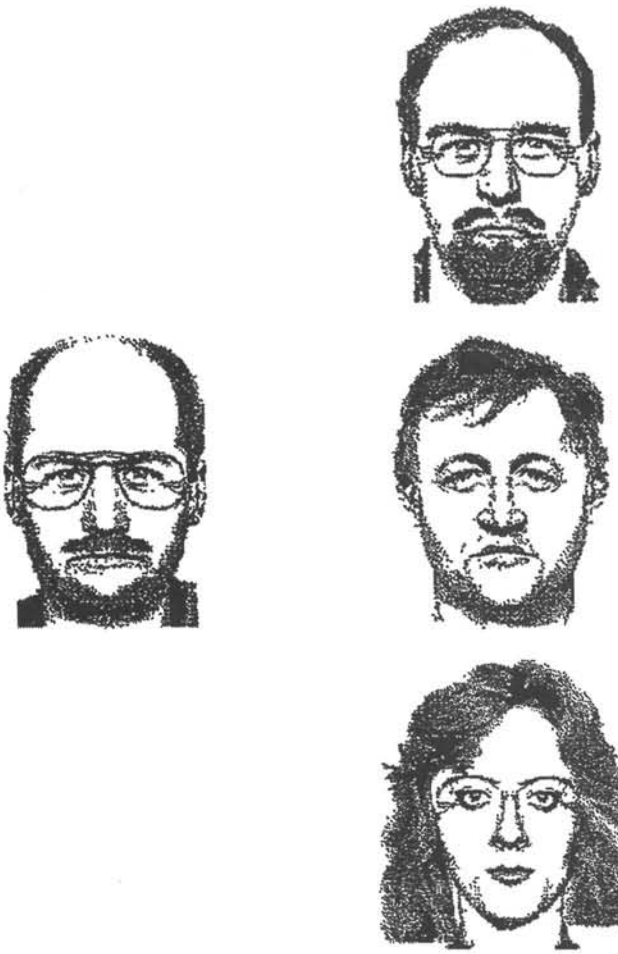

Figure 5. The faces that served as sources in the experiment. Source A, "Jack," is on the left. On the right, from top to bottom, are Sources B: "John" (high similarity), "Michael" (medium similarity), and "Susan" (low similarity).

Procedure

Participants were seated at individual computer booths and were given experimental instructions and an introduction to the detective story format presented over headphones. The instructions informed participants that their memory would be tested but did not mention the source identification test. Participants then listened to narratives about Mary from two sources, A and B. The narrative from each source was divided into three segments of approximately equal length. The segments from the two sources were interleaved in an ABABAB pattern to lessen the effectiveness of temporal information (e.g., which information was learned early) as a basis for source identification. Jack was arbitrarily designated as Source A. The order in which the sources presented their information was completely counterbalanced across subjects. While a source was speaking, his or her picture and name appeared on the computer monitor. Total presentation time was 8 min.

After both sources had completed their narratives, participants were instructed to remove their headphones. No further auditory informa-

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

tion was presented during the experiment. Instructions for the source-monitoring test were then presented on the computer screen. During the test, pictures of both sources with their names appeared side by side on the computer monitor. Fifty-two test sentences (e.g., "Mary hates lemons") were presented one at a time above the pictures. The test sentences were equally divided between items that had originally been presented by the two sources. Half of the item tests for each source were targets, and the other half were distractors. Test order was randomized for each participant. Eight practice sentences preceded the 52 test sentences. Two versions of the test stimuli were constructed such that items that were tested as targets in Version 1 were tested as distractors in Version 2, and vice versa. Test version was completely counterbalanced across subjects.

There were three response options in the self-paced memory test. Participants were asked whether the test sentence had been presented by Source A, had been presented by Source B, or had not been presented during the learning phase of the experiment ("Neither"). Each response option was indicated in a different color on the computer screen throughout the testing sequence. Participants responded by pressing color-coded keys on the computer keyboard. Whether Source A or Source B appeared on the right side of the screen during the testing sequence was completely counterbalanced across subjects. The color-coded response key for each source was always directly below the picture of the source. The response key for "Neither" was always in the middle of the keyboard.

Results

Data were initially analyzed with 3 (source similarity) × 3 (distractor similarity) ANOVAs. Separate analyses were conducted by using *d'*, HR, FAR, and ACSIM as dependent measures. The alpha level for the *F* tests was set to .05 to guarantee a fairly high level of statistical power for most of the tests. Given medium effect sizes ($f = .25$; see Cohen, 1988) and $\alpha = .05$, the power of the $F(2, 207)$ tests equalled $1 - \beta = .91$, and the power of the $F(4, 207)$ tests equalled $1 - \beta = .84$.[4]

Group means for *d'*, FAR, HR, and ACSIM are illustrated in Figures 6, 7, 8, and 9, respectively. In these figures, slopes indicate differences attributable to changes in distractor similarity; whereas the spread between lines indicates differences

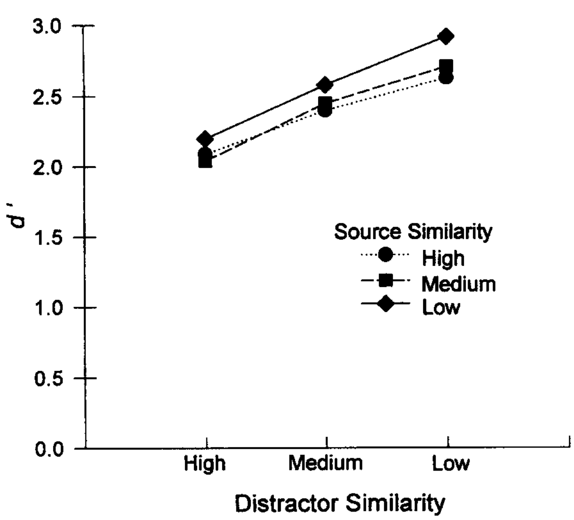

Figure 6. *d'* as a function of distractor similarity and source similarity.

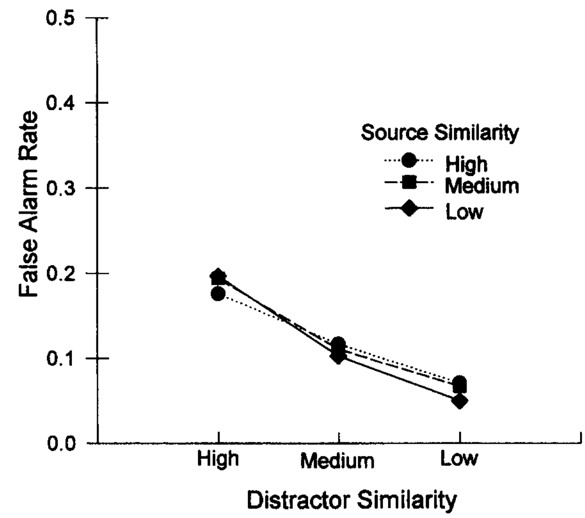

Figure 7. False-alarm rate as a function of distractor similarity and source similarity.

caused by changes in source similarity. In the following discussion, the terms *item detection* and *source discrimination* are used to refer to participant performance as measured by *d'* and by ACSIM, respectively.

The analysis using *d'* as the dependent measure was calculated on individual participant *d'*s. To avoid undefined values of *d'*, we truncated HRs greater than .96 to .96 (which occurred for 21 participants), and we increased FARs less than .04 to .04 (which occurred for 24 participants). For 3 participants, both HR and FAR were adjusted in this manner. The analysis yielded a significant main effect of distractor similarity on item detection, $F(2, 207) = 21.17$, $MSE = .354$. Examination of Figure 6 shows that the effect was in the predicted direction; increases in distractor similarity resulted in decreases in item detection. Neither the main effect of source similarity, $F(2, 207) = 2.24$, $MSE = .354$, nor the interaction between source and distractor similarity, $F(4, 207) = 0.16$, $MSE = .354$, was significant. As expected, manipulation of source similarity had no effect on item recognition.

Analyses that were based on HR and FAR confirmed that the manipulation of distractor similarity depressed item detection by increasing FAR. There was a significant main effect of distractor similarity on FAR, $F(2, 207) = 38.76$, $MSE = .008$. Examination of Figure 7 shows that this effect was in the predicted direction; FAR increased as distractor similarity increased. Neither the main effect of source similarity, $F(2, 207) = 0.14$, $MSE = .008$, nor the interaction of source and distractor similarity, $F(4, 207) = 0.41$, $MSE = .008$, was significant. Examination of Figure 8 shows that distractor similarity had no reliable effect on HR, $F(2, 207) = 0.50$, $MSE = .008$. Source similarity also had no reliable effect on HR, $F(2, 207) = 2.72$, $MSE = .008$, and did not reliably

[4] All power analyses reported in this article were computed by means of the GPOWER program (Faul & Erdfelder, 1992; see Erdfelder et al., in press).

interact with distractor similarity, $F(4, 207) = 0.34$, $MSE = .008$.

To test the robustness of the above findings and to provide a more fine-grained analysis for the purpose of testing the multinomial models, we complemented the overall two-way ANOVAs with separate one-way ANOVAs calculated at each level of source similarity using distractor similarity as the independent variable. At each level of source similarity, increases in distractor similarity produced decreases in d': high source similarity, $F(2, 69) = 4.91$, $MSE = .360$; medium source similarity, $F(2, 69) = 7.74$, $MSE = .35$; and low source similarity, $F(2, 69) = 8.91$, $MSE = .349$. At each level of source similarity, increases in distractor similarity also produced increases in FAR: high source similarity, $F(2, 69) = 9.42$, $MSE = .007$; medium source similarity, $F(2, 69) = 12.13$, $MSE = .008$; and low source similarity, $F(2, 69) = 18.00$, $MSE = .007$. As expected, changes in distractor similarity had no reliable effects on HR at any level of source similarity, all $Fs(2, 69) < .697$, all $ps > .5$.

A two-way ANOVA using ACSIM as the dependent measure yielded a significant main effect of source similarity on source discrimination, $F(2, 207) = 44.94$, $MSE = .016$. Examination of Figure 9 shows that the effect was in the predicted direction; source discrimination decreased as source similarity increased. Neither the main effect of distractor similarity, $F(2, 207) = 1.386$, $MSE = .016$, nor the interaction between source and distractor similarity, $F(4, 207) = 0.325$, $MSE = .016$, was significant. Separate one-way ANOVAs calculated at each level of distractor similarity confirmed these results. Increases in source similarity resulted in decreases in source discrimination at each level of distractor similarity: high distractor similarity, $F(2, 69) = 20.30$, $MSE = .015$; medium distractor similarity, $F(2, 69) = 13.00$, $MSE = .017$; and low distractor similarity, $F(2, 69) = 12.36$, $MSE = .014$.

The combined findings that changes in source similarity affected ACSIM and had no reliable effects on d' are difficult to reconcile with the multidimensional SDT model discussed by Thomas and Olzak (1992). According to this model, d' may provide a biased estimate of item detection because it is affected both by changes in discriminability between targets and distractors and by changes in similarity between classes of targets along the dimensions that define the decision space. The reliable effects of changes in source similarity on ACSIM indicate that increases in perceptual similarity increased the similarity between Source A and Source B. Thus, a multidimensional SDT model of source monitoring that is based on the Thomas and Olzak model would predict that d' should vary with changes in source similarity, which it did not.

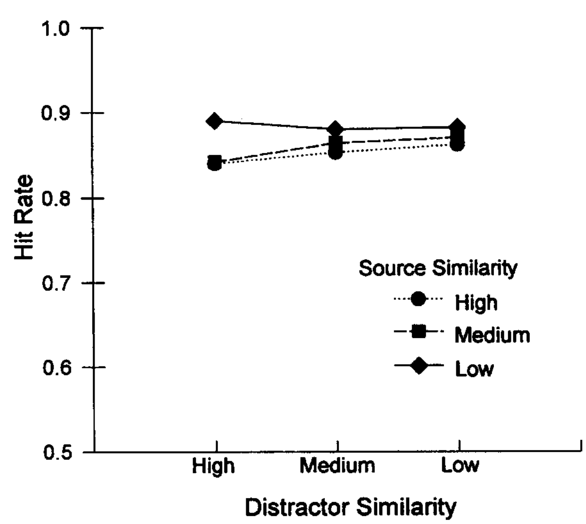

Figure 8. Hit rate as a function of distractor similarity and source similarity.

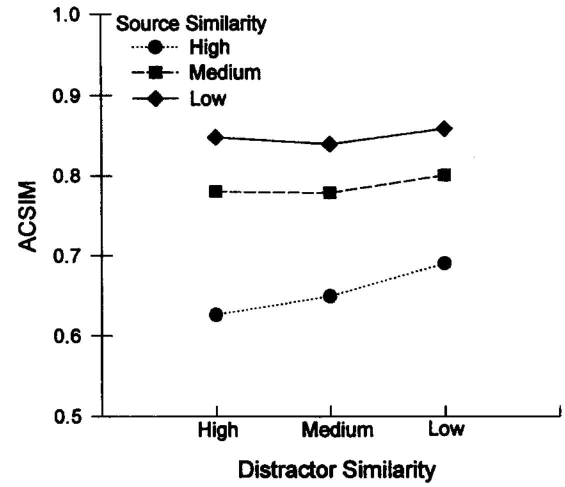

Figure 9. ACSIM (average conditional source identification measure) as a function of distractor similarity and source similarity.

In summary, the manipulations of source and distractor similarity produced the expected results. Increasing the perceptual similarity of sources decreased source discrimination and had no discernible effect on participants' ability to recognize items. Increasing the similarity of recognition distractors to studied items decreased item detection and had no significant effect on participants' ability to discriminate the source of information. Furthermore, increasing distractor similarity reduced item detection by increasing FAR and, as expected, had no reliable effect on HR.

The data from this experiment provide a rigorous test for theories of source monitoring that propose separate theoretical bases for item recognition and source discrimination because distractor similarity and source similarity had separate and independent effects across all levels of both variables. In the remainder of this article we use these data to test the three multinomial models discussed above. However, detailed theories of source monitoring are a fairly recent addition to the literature and new models are being introduced with some regularity. Because other investigators may find these data useful for testing new theories, response frequency data for the nine cells of the experimental design are given in Appendix C.

Testing the Multinomial Models

Identification of Submodels for Testing

To reiterate a point made earlier, the complete seven- and eight-parameter versions of the models illustrated in Figures

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Table 3
Four-Parameter Versions of the 1HTSM, LTSM, and 2HTSM Models

Model	Parameter D	d	b	g
1HTSM	P(CDT)	P(CDS)	P(guess old)	P(guess source is A)
LTSM	P(CDT)	P(CDS)	P(FA)	P(guess source is A)
2HTSM	P(CDT or CDD)	P(CDS)	P(guess old)	P(guess source is A)

Note. 1HTSM = one-high-threshold source-monitoring model; LTSM = low-threshold source-monitoring model; 2HTSM = two-high-threshold source-monitoring model. P(CDT) = probability of a correct detection of a target; P(CDS) = probability of a correct discrimination of a source; P(FA) = probability of a false alarm; P(CDD) = probability of a correct detection of a distractor.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

1–3 are unidentifiable. However, several identifiable submodels can be defined by imposing equality restrictions on selected subsets of parameters, as is illustrated in Figure 4. This raises the problem of determining which model from each family of models to subject to empirical test. We opted to solve this problem by identifying the submodel with the smallest number of parameters that provided an empirically adequate description of the data.

The most parsimonious submodels that seem reasonable for our experimental paradigm contain four parameters each. These models are described in Table 3. For all models it was assumed that the probabilities of correctly recognizing targets from Sources A and B are equal ($D_1 = D_2 = D$), the probabilities of correctly discriminating the source of items from Sources A and B are equal ($d_1 = d_2 = d$), and the probabilities of guessing targets or distractors came from Source A are equal ($a = g$). In addition, for the 2HTSM model it was assumed that the probabilities of detecting targets and distractors are equal ($D_1 = D_2 = D_3 = D$). This assumption was made for two reasons. First, the five-parameter version of the 2HTSM model in which D_1 is set equal to D_2 and in which D_3 is allowed to vary is unidentifiable. Second, the restriction $D_1 = D_2 = D_3$ is implied by the standard assumption made for the 2HT model of simple item detection that the probabilities of correctly identifying targets and distractors are equal. Snodgrass and Corwin (1988) have shown that 2HT models that include this assumption compare favorably with SDT models of simple item detection.

All multinomial model-based analyses reported in this article were carried out with a computer program by Hu (1991). The fit of the four-parameter versions of the three models was tested as follows. Goodness of fit was independently evaluated for each cell of the experimental design by using the log-likelihood ratio statistic G^2, which is asymptotically chi-square distributed with two degrees of freedom for the four-parameter versions of the models (see Batchelder & Riefer, 1990; Read & Cressie, 1988; Riefer & Batchelder, 1988). Values of G^2 are shown in Table 4. A significant value of G^2 indicates a poor fit of the model. Values of G^2 greater than 5.99 are significant, with $\alpha = .05$. As can be seen from the table, none of the chi-square tests reached significance. Thus, the four-parameter versions of the models provided an adequate fit to the data from all nine experimental cells.

Note that the G^2 statistics for each of the three four-parameter models are identical. This was not due to a peculiarity of our data. Equal goodness-of-fit indices will occur with some regularity whenever these four-parameter models are applied to data. In fact, the goodness of fit of the four-parameter 1HTSM model must always be equal to or better than the fit of the four-parameter 2HTSM model, and the fit of the four-parameter LTSM model must always be equal to or better than the fit of the four-parameter 1HTSM model. This can be proved by equating the model equations of two of the four-parameter models with each other and by solving for the parameters of one of the models (see Erdfelder, Murnane, & Bayen, 1995). By using this method, it can be shown that any matrix of response probabilities generated from the model equations of the 2HTSM model may also be generated from the model equations of the 1HTSM model. However, probability matrices generated from the 1HTSM model with $d_{1HT} > 1 - b_{1HT}$ cannot be expressed in terms of the 2HTSM model with parameters inside the closed interval [0, 1]. In an analogous manner, any matrix of response probabilities generated from the 1HTSM model may also be generated from the LTSM model. Probability matrices generated from the LTSM model with either $b_{LT} \geq D_{LT}$ or $d_{LT} > (1 - b_{LT}/D_{LT})/(1 - b_{LT})$ cannot be expressed in terms of the 1HTSM model with parameters in the closed interval [0, 1].

The fact that the fit of the four-parameter LTSM model to single data sets can never be worse than the fit of the other two models should not be considered an a priori advantage of the

Table 4
Results of Goodness-of-Fit Tests for the Four-Parameter Versions of the 1HTSM, LTSM, and 2HTSM Models

Distractor similarity/ source similarity	G^2
High	
High	4.09
Medium	0.84
Low	4.91
Medium	
High	0.23
Medium	0.83
Low	1.24
Low	
High	1.87
Medium	0.73
Low	1.38

Note. 1HTSM = one-high-threshold source-monitoring model; LTSM = low-threshold source-monitoring model; 2HTSM = two-high-threshold source-monitoring model. G^2 = log-likelihood ratio chi-square statistic.

Table 5
Parameter Estimates for the Four-Parameter One-High-Threshold Model of Source Monitoring for all Experimental Cells

Source similarity	Distractor similarity: High D	d	g	b	Medium D	d	g	b	Low D	d	g	b
High	.809	.290	.424	.176	.837	.314	.459	.117	.857	.391	.516	.071
Medium	.809	.592	.373	.194	.849	.587	.390	.111	.868	.609	.409	.067
Low	.872	.712	.393	.197	.871	.695	.436	.103	.884	.721	.397	.050

Note. D = item detection parameter; d = source discrimination parameter; g = probability of guessing an item is from Source A; b = probability of guessing an item is old.

model. The generality of the LTSM model is primarily due to the fact that it alone allows for the unlikely, if not impossible, case in which false-alarm rates are larger than hit rates. The 1HTSM model is more general than the 2HTSM model because the 1HTSM model, but not the 2HTSM model, allows for cases in which either the source discrimination parameter d_{1HT} is extremely high or there is a large bias b_{1HT} in favor of guessing old. Cases like these should be rare in practice. We may conclude, therefore, that deviations in the goodness of fit of the three models are possible in principle but improbable in the context of applications to typical source-monitoring experiments. Thus, although the three models embody different theoretical conceptions about the cognitive processes underlying performance in a source-monitoring paradigm, the identical fit of the three models observed in our experiment is expected.

Obviously, the goodness of fit of the models to single data sets is not very informative regarding the psychological validity of the models' parameters. For this reason empirical evaluation of the models must be based on the differential sensitivity of the model parameters to experimental variables known to affect the cognitive processes of interest, that is, item detection and source discrimination. The best fitting sets of parameters for each of the nine experimental cells for all three submodels are given in Tables 5, 6, and 7.

Overview of the Comparative Tests of the 1HTSM, LTSM, and 2HTSM Models

The comparative tests of the three four-parameter models were carried out as follows. One independent variable (e.g., distractor similarity) was varied across its three levels (high, medium, and low); whereas the other independent variable (e.g., source similarity) was fixed at each of its three levels. The fit for each model was then examined in two ways at each level of the fixed variable. First, a parameter (e.g., D, the item parameter) that is predicted to be sensitive to the varied variable (e.g., distractor similarity) was evaluated. This parameter should change in the appropriate direction to reflect the observed changes in the varied variable. Second, a parameter (e.g., d, the source parameter) that is predicted to be sensitive to the fixed variable (e.g., source similarity) was evaluated. This parameter should not be affected by changes in the varied variable because the observed item and source effects were independent. This procedure produces 18 separate hypothesis tests for each model when each of the two varied variables (distractor and source similarity) are crossed with the three levels of the fixed variable (high, medium, and low) and the three model parameters (D, d, and b). In all of these tests, the alpha level was set to .001. With N = 3,744 (3 experimental cells × 24 participants × 52 test items) and df = 2, this alpha level produces a statistical power greater than .99 (for a moderate effect size of w = .2; see Cohen, 1988).

Whether the model parameters behaved in an appropriate manner was tested as follows. For each of the 18 hypothesis tests, two versions of each model were compared. In the unrestricted version of each model, all of the parameters (D, d, g, and b) were allowed to vary. In the restricted version of each model, one of the three critical parameters (D, d, and b) was set equal across the three levels of the varied variable. The fit indices of the restricted and unrestricted models were then compared by using the log-likelihood ratio statistic G^2. Significant values of G^2 mean that the unrestricted model provided

Table 6
Parameter Estimates for the Four-Parameter Low-Threshold Model of Source Monitoring for all Experimental Cells

Source similarity	Distractor similarity: High D	d	g	b	Medium D	d	g	b	Low D	d	g	b
High	.843	.278	.424	.176	.856	.307	.459	.117	.867	.386	.516	.071
Medium	.846	.567	.373	.194	.865	.575	.390	.111	.877	.602	.409	.067
Low	.897	.692	.393	.197	.885	.685	.436	.103	.889	.716	.397	.050

Note. D = item detection parameter; d = source discrimination parameter; g = probability of guessing an item is from Source A; b = probability of a false alarm.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Table 7
Parameter Estimates for the Four-Parameter Two-High-Threshold Model of Source Monitoring for all Experimental Cells

Source similarity	Distractor similarity: High D	d	g	b	Medium D	d	g	b	Low D	d	g	b
High	.667	.352	.424	.529	.739	.355	.459	.448	.796	.420	.516	.346
Medium	.652	.735	.373	.558	.755	.660	.390	.451	.809	.653	.409	.353
Low	.700	.886	.393	.658	.782	.774	.436	.471	.840	.759	.397	.310

Note. D = item detection parameter; d = source discrimination parameter; g = probability of guessing an item is from Source A; b = probability of guessing an item is old.

the better fit. The unrestricted model should provide the better fit when the parameter that is fixed in the restricted model (e.g., D, the item parameter) is sensitive to the variable being varied (e.g., distractor similarity). The restricted and unrestricted models should not differ (and hence the G^2 statistic should not be significant) when the parameter that is fixed in the restricted model (e.g., d, the source parameter) is sensitive to the variable that is fixed (e.g., source similarity).

A valid model should produce the following pattern of results across the 18 hypothesis tests. When distractor similarity is varied and source similarity is fixed, comparison of the restricted and unrestricted models should produce significant values of G^2 at each level of source similarity for the item detection and recognition bias parameters (D and b) and nonsignificant values of G^2 at each level of source similarity for the source parameter (d). These hypotheses are implied by the assumptions that (a) distractor similarity does not affect source discrimination as measured by d, (b) distractor similarity has a negative effect on item detection as measured by D, and (c) distractor similarity has a positive effect on response bias as measured by b. The last assumption rests on the expectation that increasing the similarity of targets and distractors increases subjective estimates of the ratio of targets to distractors in the test list and hence produces an increase in bias to respond "old" (Broadbent, 1971; Buchner, Erdfelder, & Vaterrodt-Plünnecke, 1995).

Likewise, when source similarity is varied and distractor similarity is fixed, comparison of the restricted and unrestricted models should produce significant values of G^2 at each level of distractor similarity for the source parameter (d) and nonsignificant values of G^2 at each level of distractor similarity for the item detection and recognition bias parameters (D and b). These hypotheses are based on the assumptions that (a) source similarity has a negative effect on source discrimination as measured by d, and (b) source similarity has no effect on either item detection as measured by D or recognition bias as measured by b.

The results of the 18 hypothesis tests for each of the three multinomial models are shown in Tables 8, 9, and 10—all of which have the same format. The varied variable is given in the first column, the fixed variable is given in the second column, and the levels of the fixed variable are given in the third column. G^2 values are given in columns 4, 5, and 6. A model is considered valid if (a) the three G^2 values in the upper halves of columns 4 and 5 of the table (distractor similarity is varied, and an item parameter is tested) and the three G^2 values in the lower half of column 6 of the table (source similarity is varied, and the source parameter is tested) are significant and (b) the three G^2 values in the upper half of column 6 of the table (distractor similarity is varied, and the source parameter is tested) and the three G^2 values in the lower halves of columns 4 and 5 of the table (source similarity is varied and an item parameter is tested) are not significant.

The Four-Parameter One-High-Threshold Source-Monitoring Model

Results of significance tests for the 1HTSM (Batchelder & Riefer, 1990) model are given in Table 8. As expected,

Table 8
Significance Tests of the Effects of Distractor Similarity and Source Similarity on the Parameters D, b, and d of Batchelder and Riefer's (1990) One-High-Threshold Model of Source Monitoring

Varied variable	Fixed variable	Level of fixed variable	Item parameter D	Item parameter b	Source parameter d
Distractor similarity	Source similarity	High	4.25	33.36**	3.22
		Medium	6.70*	47.27**	0.21
		Low	0.49	68.35**	0.37
Source similarity	Distractor similarity	High	9.14*	1.03	64.08**
		Medium	2.69	0.66	54.35**
		Low	1.86	2.77	44.54**

Note. Values in the table are the log-likelihood ratio chi-square statistic G^2, $df = 2$. D = item detection parameter; b = probability of guessing an item is old; d = source discrimination parameter.
* $p < .05$. ** $p < .001$.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

increases in source similarity had a significant effect on the source discrimination parameter, *d*, at all three levels of the distractor similarity factor. Also as expected, changes in distractor similarity had no reliable effect on the source discrimination parameter. Thus, differences in source discrimination were accurately reflected by changes in the appropriate parameter of the 1HTSM model, and this parameter was insensitive to changes in item recognition. The 1HTSM model provided an adequate measure of source discrimination in our experiment. This result is consistent with Batchelder and Riefer's finding that the 1HTSM model produced acceptable fits to source-monitoring data from prior studies in which variables affecting source discrimination were manipulated.

The item detection parameter, *D*, failed to capture differences in item detection caused by changes in distractor similarity at all levels of source similarity. With the significance level relaxed to $\alpha = .05$, the item detection parameter was sensitive to changes in item detection at the medium level of source similarity. However, at this significance level the item detection parameter also registered a spurious difference in recognition performance when source similarity varied and distractor similarity was held constant at the high level.

The *b* parameter in the 1HTSM model measures the probability that either a distractor or an undetected target is guessed old. As expected, the *b* parameter accurately reflected the change in response bias caused by increased distractor similarity at all levels of source similarity. In addition, changes in source similarity had no reliable effects on the *b* parameter at any level of distractor similarity.

In summary, the 1HTSM model accurately measured both the changes in source discrimination caused by changes in source similarity and the change in response bias caused by increases in distractor similarity. However, the model failed to measure accurately changes in item detection caused by changes in distractor similarity.

The Four-Parameter Low-Threshold Source-Monitoring Model

Results of significance tests for the LTSM model (Batchelder, Riefer, & Hu, 1994) are given in Table 9. As was the case for the 1HTSM model, increases in source similarity had a significant effect on the source discrimination parameter, *d*, at all three levels of the distractor similarity variable. Also, as was the case with the 1HTSM model, changes in distractor similarity had no reliable effect on the source discrimination parameter. Thus, differences in source discrimination were accurately reflected by changes in the appropriate parameter of the LTSM model, and this parameter was insensitive to changes in item recognition. This result is consistent with Batchelder, Riefer, and Hu's (1994) finding that the LTSM model produced acceptable fits to prior source-monitoring studies in which variables affecting source discrimination were manipulated.

The item detection parameter, *D*, failed to capture changes in item detection caused by changes in distractor similarity at any level of source similarity. With $\alpha = .05$, the *D* parameter registered a spurious difference in item detection when source similarity varied and distractor similarity was held constant at the high level.

As we have shown elsewhere, the *b* parameters of the four-parameter 1HTSM and LTSM models are numerically identical for all possible data sets (cf. Erdfelder et al., in press). However, in the 1HTSM model, the *b* parameter measures both the probability of making a false alarm and the probability of guessing old for undetected targets, whereas in the LTSM model, the *b* parameter measures only the probability of making a false alarm. The *b* parameter in the LTSM model accurately reflected the increase in false alarms caused by increased distractor similarity at all levels of source similarity. In addition, changes in source similarity had no reliable effects on the *b* parameter at any level of distractor similarity.

As was the case for the 1HTSM model, the LTSM model accurately measured the changes in source monitoring caused by changes in source similarity. The LTSM model also accurately measured the increase in false alarms caused by increases in distractor similarity. However, as was the case for the 1HTSM model, the LTSM model failed to accurately measure the changes in item detection caused by changes in distractor similarity. Reliance on the LTSM model would lead to the mistaken conclusion that item detection was unaffected by changes in distractor similarity.

In addition to this empirical failure, the LTSM model has an important theoretical limitation. Batchelder, Riefer, and Hu's (1994) variant of the typical LTSM model does not have a

Table 9
Significance Tests of the Effects of Distractor Similarity and Source Similarity on the Parameters D, b, and d of Batchelder, Riefer, and Hu's (1994) Low-Threshold Model of Source Monitoring

Varied variable	Fixed variable	Level of fixed variable	Item parameter *D*	Item parameter *b*	Source parameter *d*
Distractor similarity	Source similarity	High	1.45	33.36**	3.48
		Medium	2.49	47.27**	0.59
		Low	0.54	68.35**	0.61
Source similarity	Distractor similarity	High	10.25*	1.03	66.10**
		Medium	2.38	0.66	55.59**
		Low	1.48	2.77	45.35**

Note. Values in the table are the log-likelihood ratio chi-square statistic G^2, $df = 2$. *D* = item detection parameter; *b* = probability of a false alarm; *d* = source discrimination parameter.
$^*p < .05$. $^{**}p < .001$.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

guessing parameter for recognition. This is an important omission when one considers that one of the important advantages of the multinomial approach to source monitoring is that these models provide independent estimates of item detection, source discrimination, and response biases. Giving up the capacity to provide an independent measure of response bias in recognition weakens the model.

The Four-Parameter Two-High-Threshold Source-Monitoring Model

Results of significance tests for the 2HTSM model are given in Table 10. As was the case for both the 1HTSM and LTSM models, increases in source similarity had a significant effect on the source discrimination parameter, *d*, at all three levels of the item variable. Also, as was the case with both the 1HTSM and LTSM models, changes in distractor similarity had no reliable effect on the source discrimination parameter. Thus, differences in source discrimination were accurately reflected by changes in the appropriate parameter of the 2HTSM model, and this parameter was insensitive to changes in item recognition.

As expected, increases in distractor similarity had a significant effect on the item detection parameter, *D*, at all levels of source similarity. Changes in source similarity had no reliable effect on the item detection parameter at any level of distractor similarity. This pattern of results did not change with $\alpha = .05$. Thus, differences in item detection were accurately reflected by changes in the appropriate parameter of the 2HTSM model, and this parameter was unaffected by changes in source discrimination.

The *b* parameter in the 2HTSM model measures the probability that either an undetected distractor or an undetected target is guessed old. The *b* parameter accurately reflected the increase in response bias caused by increased distractor similarity at all levels of source similarity. In addition, changes in source similarity had no reliable effects on the *b* parameter at the medium and low levels of distractor similarity and had an effect at the high level of distractor similarity only with the significance level relaxed to $\alpha = .05$.

In summary, the 2HTSM model accurately measured changes in source discrimination caused by changes in source similarity and accurately measured both the changes in item detection and response bias caused by changes in distractor similarity. Unlike the LTSM model, the 2HTSM model provides an independent measure of recognition bias.

General Discussion

Multinomial models are powerful theories of the decision processes involved in source monitoring. One of the valuable features of these theories is that they provide several independent parameters for measuring performance in a source-monitoring task. A typical source-monitoring task requires participants to both detect the difference between old and new test items and discriminate between sources. Accurate measurement of performance in source-monitoring paradigms demands measurement techniques that do not confound memory for the item with memory for the source and that are capable of measuring sensitivity independent of bias in both the item detection and source discrimination components of the task. No empirical measures have yet been identified that satisfy all of these demands. All of the commonly used empirical measures of source memory confound item detection and source discrimination in most circumstances, and even the rarely used ACSIM confounds item recognition and source discrimination in some circumstances. Moreover, none of the empirical measures are capable of separating sensitivity from bias in measuring source discrimination in all circumstances (Murnane & Bayen, in press). In contrast, all of the multinomial models considered in this article provide independent parameters for measuring item detection, source discrimination, and several types of response bias. If it is only source discrimination that one is interested in measuring with a source-monitoring paradigm, any one of the three multinomial models examined here is preferable to all of the empirical measures because all three models adequately measure source discrimination independent of item detection.

Of course, as theories of the decision processes involved in source monitoring, multinomial models must do more than provide a parameter that adequately measures memory for the source independent of memory for the item. The development of these powerful models introduces the need for equally powerful empirical tests of the models as theories and not just as measurement devices for source discrimination. Most source-monitoring experiments are designed to examine factors that

Table 10
Significance Tests of the Effects of Distractor Similarity and Source Similarity on the Parameters D, b, and d of the Two-High-Threshold Model of Source Monitoring

Varied variable	Fixed variable	Level of fixed variable	Item parameter *D*	Item parameter *b*	Source parameter *d*
Distractor similarity	Source similarity	High	23.06**	12.31*	1.38
		Medium	34.73**	15.56**	1.95
		Low	30.99**	38.06**	5.85
Source similarity	Distractor similarity	High	2.84	8.71*	61.71**
		Medium	2.83	.20	49.03**
		Low	3.78	.56	39.84**

Note. Values in the table are the log-likelihood ratio chi-square statistic G^2, $df = 2$. D = item detection parameter; b = probability of guessing an item is old; d = source discrimination parameter.
$^*p < .05$. $^{**}p < .001$.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

affect source discrimination; factors affecting item detection or bias are either ignored or experimentally controlled. Thus, the results of these experiments do not provide an adequate test of the multinomial models as theories of the decision processes involved in source monitoring.

Evaluating the three multinomial models against the results of our experiment illustrates the importance of testing the ability of theories of source monitoring to measure accurately both memory for the item and memory for the source. The one-high-threshold source-monitoring model (1HTSM) developed by Batchelder and Riefer (1990) and the low-threshold model of source monitoring (LTSM) developed by Batchelder, Riefer, and Hu (1994) accurately measured both changes in source discrimination when they occurred and increases in response bias that occurred when distractor similarity increased. However, both models also failed to capture changes in item detection. The LTSM model has the added limitation that it is incapable of providing separate and independent estimates of guessing in recognition. Only the two-high-threshold source-monitoring model (2HTSM) provided accurate measures of both item detection and source discrimination across all nine cells of the experimental design, captured the change in response bias when distractor similarity increased, and retained the ability to measure biases in both item detection and source discrimination.

Our discussion of theories of source monitoring thus far has not mentioned any of the well-developed formal theories of memory that are of current interest (for reviews, see Hintzman, 1990; Raaijmakers & Shiffrin, 1992). Strictly speaking, multinomial models are theories of decision processes, not memory processes. To the best of our knowledge, formal models of the complex memory processes involved in source monitoring have yet to be developed. It seems plausible that powerful theories of recognition such as Minerva 2 (Hintzman, 1988), SAM (Gillund & Shiffrin, 1984), or the general context model (Murnane & Phelps, 1994) may be able to be adapted to model the memory component of a source-monitoring task. However, a detailed discussion of possible formal memory models of source monitoring is beyond the scope of this article.

Multinomial theories of source monitoring are variants within the general class of threshold theories of decision. However, most current theories of recognition make use of a model of decision that is derived from SDT. The relative value of SDT and one-high-threshold models of the decision component of a source-monitoring task has been discussed by Kinchla (1994) and Batchelder, Riefer, and Hu (1994). When the three models were introduced, we pointed out that although the one-high-threshold model for simple item detection is widely agreed to be inferior to SDT, both low-threshold and two-high-threshold models of simple item detection fare very well when compared with models derived from SDT (Macmillan & Creelman, 1990; Macmillan & Kaplan, 1985; Snodgrass & Corwin, 1988; Swets, 1986). Moreover, it is unclear whether the criticisms that can be brought to bear against one-high-threshold models of simple detection carry equal weight when brought to bear against threshold models of decision in the more complex decision environment imposed by the typical source-monitoring task. Finally, recent evidence indicates that the homogeneous variance assumption that underlies common methods for calculating d' is often violated in typical recognition studies (Gronlund & Elam, 1994; Ratcliff et al., 1992). The implications of these findings for memory theories that incorporate signal detection models of decision are just beginning to be examined. Although signal detection models are more familiar to most investigators, it is far from clear whether signal detection or threshold theories will provide the best framework for modeling source monitoring.

It is clear, however, that more work remains to be done. Factors affecting bias in the source discrimination component of the task have not systematically been manipulated, and thus it remains to be seen whether the 2HTSM model is capable of accurately measuring this factor as well. In addition, the adequacy of the 2HTSM model must be shown over a variety of source-monitoring paradigms before it can be recommended as a general purpose model of source monitoring. Finally, our conclusions regarding the validity of the 1HTSM, LTSM, and 2HTSM models are confined to the four-parameter versions of these models that were found to be adequate for our experimental paradigm. It remains to be seen whether five- or six-parameter versions of the 2HTSM model are superior to corresponding versions of the other two models when the four-parameter versions of the models are inadequate (e.g., when item detection or source discrimination probabilities differ between sources).

It appears to be the case that the construction of mathematical models of source monitoring is entering a fruitful stage of development, and we eagerly await the development of SDT-based theories as well as further theory development within the multinomial framework. The pattern of findings in the experiment described in this article provides a challenge to new source-monitoring theories as they are developed. Until new source-monitoring theories are constructed that are shown to be capable of measuring the independent effects on item detection and source discrimination evident in our data, we believe the 2HTSM model is the best available model of source monitoring.

References

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

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

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

Begg, I. M., Anas, A., & Farinacci, S. (1992). Dissociation of processes in belief: Source recollection, statement familiarity, and the illusion of truth. *Journal of Experimental Psychology: General, 121,* 446–458.

Broadbent, D. E. (1971). *Decision and stress.* London: Academic Press.

Brown, A. S., & Halliday, H. E. (1991). Cryptomnesia and source memory difficulties. *American Journal of Psychology, 104,* 475–490.

Buchner, A., Erdfelder, E., & Vaterrodt-Plünnecke, B. (1995). Toward unbiased measurement of conscious and unconscious memory processes within the process dissociation framework. *Journal of Experimental Psychology: General, 124,* 137–160.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Erdfelder, E., Faul, F., & Buchner, A. (in press). GPOWER: A general power analysis program. *Behavior Research Methods, Instruments, & Computers.*

Erdfelder, E., Murnane, K., & Bayen, U. J. (1995). Die Messung kognitiver Prozesse im Paradigma der Quellendiskrimination [The measurement of cognitive processes in the source-monitoring paradigm]. In K. Pawlik (Ed.), *Bericht über den 39. Kongreß der Deutschen Gesellschaft für Psychologie in Hamburg 1994* (pp. 541–547). Göttingen, Germany: Hogrefe.

Faul, F., & Erdfelder, E. (1992). GPOWER: A priori, post-hoc, and compromise power analyses for MS-DOS [Computer program]. Bonn, Germany: University of Bonn.

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

Gerard, L. D., & Scarborough, D. L. (1989). Language-specific lexical access of homographs by bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 305–315.

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

Gopnik, A., & Graf, P. (1988). Knowing how you know: Young children's ability to identify and remember the sources of their beliefs. *Child Development, 59,* 1366–1371.

Green, D. M., & Birdsall, T. G. (1978). Detection and recognition. *Psychological Review, 85,* 192–206.

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

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

Hashtroudi, S., Johnson, M. K., Vnek, N., & Ferguson, S. A. (1994). Aging and the effects of affective and factual focus on source monitoring and recall. *Psychology and Aging, 9,* 160–170.

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

Hintzman, D. L. (1990). Human learning and memory: Connections and dissociations. *Annual Review of Psychology, 41,* 109–139.

Hintzman, D. L. (1991). Why are formal models useful in psychology? In W. E. Hockley, & S. Lewandowsky (Eds.), *Relating theory and data: Essays on human memory in honor of Bennet B. Murdock* (pp. 331–346). Hillsdale, NJ: Erlbaum.

Hu, X. (1991). Statistical inference program for multinomial binary tree models (Version 1.0) [Computer program]. Irvine, CA: University of California.

Hu, X., & Batchelder, W. H. (1994). The statistical analysis of general processing tree models with the EM algorithm. *Psychometrika, 59,* 21–47.

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

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

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

Kintsch, W. (1970). *Learning, memory, and conceptual processes.* New York: Wiley.

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

Lindsay, D. S. (1990). Misleading suggestions can impair eyewitnesses' ability to remember event details. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 1077–1083.

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

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

Luce, R. D. (1963a). Detection and recognition. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (Vol. 1, pp. 103–189). New York: Wiley.

Luce, R. D. (1963b). A threshold theory for simple detection experiments. *Psychological Review, 70,* 61–79.

Macmillan, N. A., & Creelman, C. D. (1990). Response bias: Characteristics of detection theory, threshold theory, and "nonparametric" indexes. *Psychological Bulletin, 107,* 401–413.

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

Macmillan, N. A., & Kaplan, H. L. (1985). Detection theory analysis of group data: Estimating sensitivity from average hit and false-alarm rates. *Psychological Bulletin, 98,* 185–199.

Marsh, R. L., & Bower, G. H. (1993). Eliciting cryptomnesia: Unconscious plagiarism in a puzzle task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 673–688.

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

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

Murnane, K., & Bayen, U. J. (in press). An evaluation of empirical measures of source monitoring. *Memory & Cognition.*

Murnane, K., & Phelps, M. P. (1994). When does a different environmental context make a difference in recognition? A global activation model. *Memory & Cognition, 22,* 584–590.

Murnane, K., & Shiffrin, R. M. (1991). Word repetitions in sentence recognition. *Memory & Cognition, 19,* 119–130.

Raaijmakers, J. G., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review, 88,* 93–134.

Raaijmakers, J. G. W., & Shiffrin, R. M. (1992). Models for recall and recognition. *Annual Review of Psychology, 43,* 205–234.

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

Read, T. R. C., & Cressie, N. A. C. (1988). *Goodness-of-fit statistics for discrete multivariate data.* New York: Springer-Verlag.

Riefer, D. M., & Batchelder, W. H. (1988). Multinomial modeling and the measurement of cognitive processes. *Psychological Review, 95,* 318–339.

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

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

Spencer, W. D., & Raz, N. (1994). Memory for facts, source, and context: Can frontal lobe dysfunction explain age-related differences? *Psychology and Aging, 9,* 149–159.

Swets, J. A. (1986). Indices of discrimination or diagnostic accuracy: Their ROCs and implied models. *Psychological Bulletin, 99,* 100–117.

Taylor, A. E., Saint-Cyr, J. A., & Lang, A. E. (1990). Memory and learning in early Parkinson's disease: Evidence for a "frontal lobe syndrome." *Brain and Cognition, 13,* 211–232.

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

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

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Appendix A

Narrative Statements of Source A and Source B

This appendix contains the narrative statements presented by Source A and Source B. Underlined words are critical items that were used as the basis for the recognition tests. Underlining is used to facilitate identification of critical items for the reader; the critical items were not stressed in the presentation of the narratives to participants. The segments from the two sources were interleaved during presentation in such a way that participants heard Segment 1 from Source A, followed by Segment 1 from Source B, followed by Segment 2 from Source A, and so forth.

Source A

Segment 1

My name is Jack. I am Mary's brother. I am so glad you are willing to help us. I cannot understand what happened. Everything was so normal on that day when Mary disappeared. That was last Sunday. Mary got up early in the morning and did her sit-ups. She does that every morning. When I came into the kitchen, Mary had already finished her exercise and was preparing pancakes for all of us. We talked about Jim, Mary's son from her first marriage who is now going to College. She had received a letter from him, saying that he is now working for McDonald's because he needs more money for College. He is planning to study physics. Mary wondered whether she should send him some money. She originally wanted to send him a fancy necktie but now she thought he might prefer some cash instead. Mary and I were both laughing when we talked about her son because she remembered that when he was still living here, he used to play trumpet every Sunday morning, and she would complain about the noise. She liked it better when her son was fishing which is his favorite pastime.

Segment 2

When we all had breakfast together on that day, Mary's uncle called and asked whether she would like to come over and play checkers. That is Mary's favorite game. She didn't go, however, because she had already planned to go and play racquetball later that morning. She recently joined a club, because her physician said a little more exercise would be good for her. She's the president of a small company and sits in an office all day long and doesn't get much exercise. Her favorite hobby is knitting, and you don't get a lot of exercise doing this either. Anyway, I went along with her to this club and when we came back we each had a large glass of Pepsi. Mary told me her plans for the Christmas party. She was born in North Dakota. Some of the family members are still living there, and she wanted to invite them to the party. She said that the exchange student from Sweden that she is expecting would also be here at that time. She didn't want to invite her brother Tom, though, because she hates cigarettes and he always brings those with him. We had a quick lunch, some spaghetti which was left over from the day before. After lunch, Mary had some cake, but not too much because she's concerned about her weight. She gained some pounds during her vacation in June. She probably ate too much roast beef which is Mary's favorite food.

Segment 3

In the afternoon of the Sunday when she disappeared, Mary rested for a while. Then she tried out the cassette recorder which she had bought the day before. It worked fine and she was very happy about it.

Then Mary's little nephew came to visit us with some of his toys. He's five years old and an adorable child. He said he had started collecting stamps, and asked us whether we had any for him. Well, we didn't, but we said we would be watching for some.

When her nephew left, Mary told me that she wanted to run some errands. It was late afternoon when she left. She wanted to drive to the mall and buy a book that she'd seen on sale, science fiction, those are her favorite ones. She also had to go to the pet store to get some food for her husky. I remember that she had her expensive new necklace with her when she left. She got it for her birthday. She never took it anywhere before because she was so afraid she'd lose it. So she left, and I haven't seen her since then. That's really all I can remember.

Source B

Segment 1

I am John (Michael, Susan). I am Mary's brother (husband, friend). Something must have happened to her. She wouldn't just leave without telling us. She has been living in this city for almost twenty years, and nothing like this has ever happened before. When Mary didn't come back last Sunday night, we first called her grandfather. She visits him almost every Sunday, but she wasn't there. Then we worried about an accident. Mary was driving the new Cadillac, and she was not too familiar with it yet.

But let me start at the beginning of the day. Right after breakfast, Mary left for a while. I'm not sure where she went. When she came back, she brought the latest issue of Newsweek and a couple of catalogues. In one of the catalogues, there was a bed she really liked, and she said she wanted to buy it. The other one was a vacation catalogue. Mary likes to travel. She spent her last vacation in Jamaica and she really liked it. She said she would like to take another couple of days off and go canoeing.

Segment 2

We somehow got to talking about her brother Tom who is a doctor. She complained about his bad habits and that he drinks too much. Last time he came over, he drank a lot of whiskey. Mary's afraid that her son Jim, from her first marriage, might turn out just like his uncle. Jim lives in New York and she only hears from him every two months or so. Mary received a letter from him telling her that he had moved into a dormitory. He also sent her flowers. He always does that when he needs money because he knows how to win his mother's heart. Mary loves flowers, and plants, and trees and all that. Have you seen the beautiful oak tree in the yard? She planted it six years ago. Mary loves animals, too, especially birds. Parakeets are her favorite ones.

But let me get back to what else happened that day when she disappeared. Oh yes, after lunch Mary made a meal plan for the whole week. She wrote a grocery list with her pencil. I can show you the list. She needed to buy string beans. I thought broccoli would be good, but Mary hates it and said we wouldn't have any as long as she does the shopping and cooking. She put lemons on the shopping list, because of the vitamins, she said.

Segment 3

Mary usually has a cup of tea with sugar on Sunday afternoons, but last Sunday she didn't. She took off her loafers, lay down for a while and listened to Bach, her favorite composer. She got up when her little nephew stopped by to show us his new bicycle. I wondered about his visit because last week he had the flu and I didn't expect such a fast

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

recovery. Right after that, Mary said she wanted to go and run some errands at the mall. On the way there she wanted to visit an art gallery and look at some watercolors. Mary is interested in that sort of thing.

I'll show you a picture of Mary in a minute. She has dark eyes and black hair. She is of Scottish descent. She was wearing blue slacks and a yellow blouse when she left. I remember that because she said she had to change later to go to the opera. When Mary was about to leave I realized she had left her Visa Card on the coffee table. She always carries it, so I ran after her and gave it to her. She thanked me and said she would be back in about two hours. That was the last time I saw her.

Appendix B

Item Quadruplets Used to Construct Narrative Statements and Test Items

	Distractor similarity		
Target	High	Medium	Low
Source A			
does sit-ups	does pushups	lifts weights	swims laps
pancakes	waffles	bagels	eggs
McDonald's	Burger King	Pizza Hut	JC Penney
physics	chemistry	history	car repair
necktie	bow tie	scarf	sweater
trumpet	trombone	piano	bongo drums
fishing	hunting	hiking	woodworking
uncle	aunt	niece	neighbor
checkers	chess	bridge	shuffleboard
racquetball	squash	tennis	golf
president	vice President	manager	receptionist
knitting	crocheting	sewing	gardening
Pepsi	Coke	7-Up	milk
Christmas	Thanksgiving	4th of July	birthday
North Dakota	South Dakota	Nebraska	Massachusetts
Sweden	Denmark	France	Madagascar
cigarettes	cigars	pipes	chewing gum
spaghetti	linguini	pizza	tacos
cake	pie	cookies	fruit cocktail
in June	in July	in December	on her last wedding anniversary
roast beef	steak	ham	oatmeal
cassette recorder	CD player	television	vacuum cleaner
stamps	coins	comic books	butterflies
science fiction	fantasy	mystery	poetry
husky	German shepherd	poodle	siamese
necklace	bracelet	sapphire ring	camera
Source B			
city	town	village	farm
grandfather	grandmother	cousin	friend
Cadillac	Lincoln	Toyota	motorcycle
Newsweek	Time Magazine	Cosmopolitan	The National Enquirer
bed	couch	table	bird-bath
Jamaica	Bermuda	Florida	Paris
canoeing	row-boating	water skiing	camping
doctor	dentist	lawyer	plumber
whiskey	bourbon	beer	seltzer
New York	New Jersey	Ohio	Hawaii
oak tree	maple tree	palm tree	rose bush
parakeets	canaries	robins	eagles
pencil	pen	crayon	typewriter
string beans	lima beans	peas	hamburger
broccoli	cauliflower	beets	yogurt
lemons	limes	oranges	bananas
sugar	honey	milk	her friend
loafers	sneakers	boots	gloves
bicycle	tricycle	wagon	teddy bear
the flu	a cold	measles	a sprained ankle
watercolors	oil paintings	photography	ceramics
black	brown	blond	curly
Scottish	English	Italian	Chinese
slacks	jeans	skirt	hat
opera	musical	movie	restaurant
Visa Card	Master Card	personal checks	hairbrush

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Appendix C

Response Frequencies for the Nine Cells of the Experiment

Source similarity	Distractor similarity: High "A"	"B"	"N"	Medium "A"	"B"	"N"	Low "A"	"B"	"N"
High									
A	160	112	40	166	99	47	193	83	36
B	79	175	58	86	183	43	83	182	47
N	45	65	514	33	40	551	24	20	580
Medium									
A	196	70	46	203	68	41	212	65	35
B	45	217	50	47	222	43	44	226	42
N	42	79	503	24	45	555	17	25	582
Low									
A	236	52	24	226	52	34	234	46	32
B	35	237	40	35	239	38	33	242	37
N	46	77	501	31	33	560	10	21	593

Note. "A" = "Source A" response; "B" = "Source B" response; "N" = "Neither" response. A = Source A item; B = Source B item; N = Distractor item.

Received May 16, 1994
Revision received November 30, 1994
Accepted January 10, 1995 ■

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.