Available online at www.sciencedirect.com

SCIENCE DIRECT

Journal of Mathematical Psychology 47 (2003) 292–303

Journal of Mathematical Psychology

http://www.elsevier.com/locate/jmp

Source monitoring and multivariate signal detection theory, with a model for selection☆

Lawrence T. DeCarlo*

Department of Human Development, Teachers College, Columbia University, 525 West 120th Street, Box 118, New York, NY 10027, USA

Received 2 October 2001; revised 15 January 2003

Abstract

Participants in source monitoring studies, in addition to determining whether an item is old or new, also discriminate the source of the item, such as whether the item was presented in a male or female voice. This article shows how to apply multivariate signal detection theory (SDT) to source monitoring. An interesting aspect of one version of the source monitoring procedure, from the perspective of multivariate SDT, is that it involves a type of selection, in that a discrimination response is observed only if the detection decision is that an item is old. If the selection is ignored, then the estimate of the discrimination parameter can be biased; the nature and magnitude of the bias are illustrated. A bivariate signal detection model that recognizes selection is presented and its application is illustrated. The approach to source monitoring via multivariate SDT provides new results that are informative about underlying psychological processes.

1. Introduction

Source monitoring studies have been widely used in psychology to study memory, language, and other psychological processes (for references, see Hilford, Glanzer, Kim, & DeCarlo, 2002; Johnson, Hashtroudi, & Lindsay, 1993). The source monitoring task adds another component to item recognition in that, in addition to determining whether an item (e.g., a word) was presented earlier, participants must also determine the source of the item (e.g., whether the item had been presented by source A or source B). Thus, the task consists of both detection of items and discrimination of their source.[1]

One approach to source monitoring is through multinomial models, following the presentation by Batchelder and Riefer (1990) of a multinomial model for source monitoring. Batchelder and Riefer (1999) recently provided an extensive review of applications of multinomial models to source monitoring and other areas.

Another approach to source monitoring is through multivariate signal detection theory[2] (SDT; see Ashby, 1992; Macmillan & Creelman, 1991; Tanner, 1956; Wickens, 1992). Although the possibility of applying multivariate SDT to source monitoring has previously been noted (e.g., Kinchla, 1994), the approach has not been developed in any detail nor has it been empirically investigated. This is somewhat unfortunate because a number of recent studies have used a multivariate procedure to investigate source monitoring, in that participants gave both detection and discrimination responses (e.g., Hilford et al., 2002; Mather, Johnson, & De Leonardis, 1999; Slotnick, Klein, Dodson, & Shimamura, 2000; Yonelinas, 1999), yet none of these studies have used a multivariate SDT analysis, nor was the multivariate nature of the data (i.e., two responses were given on each trial) recognized in the analysis.

This article discusses source monitoring from the perspective of multivariate SDT. It is shown that

☆Programs for some of the examples presented here are available at the authors website, http://www.columbia.edu/~1d208.

*Fax: 212-678-3837.

E-mail address: decarlo@exchange.tc.columbia.edu.

[1] The term identification has also been used to describe the task; the focus here is on situations where the participant simply discriminates between two sources, A and B, and so the term discrimination is used.

[2] Multivariate SDT refers to situations where the underlying stimuli are multidimensional and two or more responses are given, typically with at least one response for each dimension. Multidimensional SDT involves multidimensional stimuli but not necessarily two or more responses.

0022-2496/03/$ - see front matter © 2003 Elsevier Science (USA). All rights reserved.
doi:10.1016/S0022-2496(03)00005-1

previous studies have ignored information in the data by in essence treating it in a univariate manner. An analysis via multivariate SDT, on the other hand, is shown to offer a concise summary of the data; it also provides new results that are informative about underlying psychological processes (e.g., estimates of the bivariate correlations of the underlying distributions). The implications of multivariate SDT for the analysis of two basic versions of the source monitoring procedure are discussed. Several studies, for example, have used a procedure where participants gave both a detection and discrimination response on each trial; the application of multivariate SDT in this case is straightforward and examples are presented in the next section. In other studies, a *conditional* source monitoring procedure was used, in that participants gave a discrimination response only if the detection response was "old". This version of the procedure involves a kind of response-selection, in that the discrimination response is not observed when the detection response is "new". It is shown that, in order to apply multivariate SDT to this situation, the effects of selection must be taken into account. A formal multivariate SDT model that recognizes selection is presented. An application of the model with selection to source monitoring data shows that the results are consistent with those obtained for the procedure that does not involve selection. In addition, it is shown that ignoring the effects of selection can give misleading results.

The remainder of the article is organized as follows. First, a bivariate normal SDT model with rating responses is presented and applied to data from several source monitoring studies. Next, a bivariate SDT model for the situation involving response selection is developed and applied to source monitoring data. It is shown that ignoring the effects of selection can bias the estimate of the discrimination parameter. The last section discusses implications of multivariate SDT for some commonly used univariate versions of the source monitoring procedure.

2. Source monitoring and multivariate signal detection theory

2.1. Bivariate SDT with rating responses

As noted above, several recent studies have required participants to give separate detection and discrimination decisions, with rating responses used for each decision (e.g., Slotnick et al., 2000; Yonelinas, 1999).[3] The analyses in these articles, however, treated the data in a univariate manner, in that a possible correlation between the detection and discrimination responses was not considered and discrimination was simply estimated from the marginal frequencies, or marginal ROC curves were examined. In the present approach, multivariate SDT is used and estimates of the detection, discrimination, and criteria parameters and their standard errors are obtained, along with estimates of the standard deviations and correlations of the underlying distributions.

To start, note that from the perspective of multivariate SDT as applied to source monitoring, the effects of a presentation of each item (new, A, or B) can be represented by underlying bivariate probability distributions (i.e., three bivariate distributions for three items). The bivariate distributions represent multidimensional perceptions of the stimuli. That is, for source monitoring, the stimuli differ not only on a dimension that is used for the detection decision, but also on a dimension that is used for the discrimination decision. The detection decision is usually thought of as being based on a dimension of familiarity whereas the discrimination decision can be thought of as being based on a dimension of features (that are used to discriminate the stimuli). Note that the same approach has been used in studies involving simultaneous detection and identification (see Ashby (1992), and the chapters therein; Macmillan & Creelman, 1991; Tanner, 1956).

More formally, let the latent variables used for the decisions be denoted as y^*_{jg}, with $j = 1, 2$ indicating the two dimensions (e.g., familiarity and features, respectively) and $g = \mathrm{N, A, B}$ indicating the three items (where $\mathrm{N} = $ new). With respect to multivariate SDT, what is referred to in structural equation modeling (SEM; e.g., Bollen, 1989) as the structural part of the model in this case is simply a model for the means, variances, and covariances of the three bivariate distributions associated with new, A, and B items. Specifically, the structural model is

$$y^*_{jg} = \Psi_{jg} + \varepsilon_{jg}, \tag{1}$$

where E is the expectation operator and $E(\varepsilon_{jg}) = 0$. It follows that $E(y^*_{jg}) = \Psi_{jg}$, where Ψ_{jg} are the means of the underlying bivariate distributions on dimension j for item g. The variances of the underlying distributions are denoted here as $V(\varepsilon_{jg}) = \sigma^2_{jg}$ and the bivariate correlations as $\mathrm{corr}(\varepsilon_{1g}, \varepsilon_{2g}) = \rho_g$.

A second basic assumption of multivariate SDT is that observers use response criteria on each dimension to divide the space into (confidence) decisions of new or old, for detection, or A or B, for discrimination. Let the observed response variable be denoted as Y_{jg} and let $k_j (k_j = 1, 2, \ldots, K_j)$ indicate K_j ordered response categories for dimension j (the subscript j is used on K_j because the number of response categories do not have

[3] Banks (2000) also conducted an experiment where both detection and discrimination responses were obtained, however the detection responses were all given first followed by the discrimination responses, so the procedure did not involve obtaining two responses on each trial. Nevertheless, a multivariate SDT model could be applied to the data.

to be the same across the two decisions). The decision rule can then be written as

$$Y_{jg} = k_j \quad \text{if} \quad c_{j,k-1} < y^*_{jg} \leqslant c_{jk}, \tag{2}$$

where $c_{j0} = -\infty$, $c_{jK} = \infty$, and $c_{j1} < c_{j2} < \cdots < c_{j,K-1}$ (note that the subscript j is not needed on k in c_{jk} because the dimension is clear; also the subscript g is not needed on c_{jk} because the criteria are assumed to have the same location across the g items). Eq. (2) reflects an assumption referred to as *decision separability* by Ashby and Townsend (1986).

It follows directly from Eqs. (1) and (2) that

$$\begin{aligned} p(Y_{1g} \leqslant k_1, Y_{2g} \leqslant k_2) &= p(y^*_{1g} \leqslant c_{1k}, y^*_{2g} \leqslant c_{2k}) \\ &= p(\varepsilon_{1g} \leqslant c_{1k} - \Psi_{1g}, \varepsilon_{2g} \leqslant c_{2k} - \Psi_{2g}). \end{aligned}$$

If $(\varepsilon_{1g}, \varepsilon_{2g})$ are bivariate normal with variances σ^2_{jg} and bivariate correlations ρ_g then

$$\begin{aligned} &p(\varepsilon_{1g} \leqslant c_{1k} - \Psi_{1g}, \varepsilon_{2g} \leqslant c_{2k} - \Psi_{2g}) \\ &\quad = \Phi_2[(c_{1k} - \Psi_{1g})/\sigma_{1g}, (c_{2k} - \Psi_{2g})/\sigma_{2g}, \rho_g], \end{aligned}$$

where Φ_2 is the bivariate normal cumulative distribution function (CDF). The result is a bivariate normal SDT model for rating responses,

$$\begin{aligned} &p(Y_{1g} \leqslant k_1, Y_{2g} \leqslant k_2) \\ &\quad = \Phi_2[(c_{1k} - \Psi_{1g})/\sigma_{1g}, (c_{2k} - \Psi_{2g})/\sigma_{2g}, \rho_g], \end{aligned} \tag{3}$$

for g = new, A, and B.

The bivariate normal SDT model of Eq. (3) is identified for a source monitoring study with two sources, that is, a unique solution for the parameters exists, if rating responses with at least three categories are used in each component (the use of only two response categories is discussed below). As in univariate SDT, one of the distributions is used as a reference and so its mean and standard deviation on each dimension are set to zero and unity, respectively. The new item distribution was used here as the reference, and so $\Psi_{jN} = 0$ and $\sigma_{jN} = 1$ for $j = 1, 2$. With this parameterization, the means of the A and B distributions on each dimension give the detection parameters $d_A = \Psi_{1A}$ and $d_B = \Psi_{1B}$ and the discrimination parameter $d_{AB} = \Psi_{2B} - \Psi_{2A}$, which is simply the distance between the means of distributions A and B on dimension 2. Fig. 1 illustrates the theory; the figure shows contours for three bivariate normal distributions, corresponding to new, A, and B items, and two response criteria on each dimension, which delineate responses of 1–3.

Wickens (1992) provides details about full information maximum likelihood (FIML) estimation for the multivariate SDT model. The (log) likelihood function to be maximized follows directly from Eq. (3), with the probabilities for each response pattern obtained by subtracting the bivariate cumulative probabilities. More generally, Eq. (3) can be viewed as a multiple-group bivariate probit model with ordinal indicators (with an

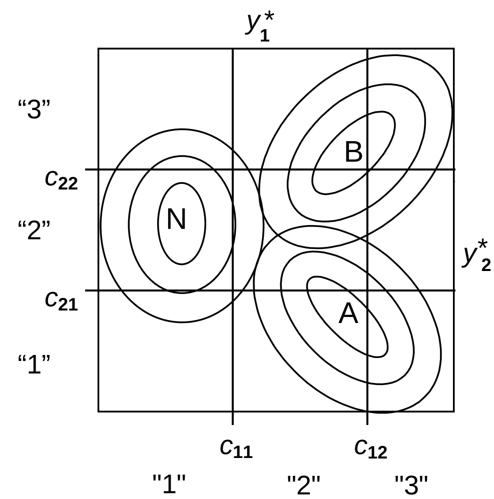

Fig. 1. Contours for new, A, and B items for a bivariate normal SDT model applied to source monitoring. The familiarity dimension is y^*_1, the feature dimension is y^*_2; c_{11} and c_{12} are response criteria on dimension 1; c_{21} and c_{22} are response criteria on dimension 2.

intercept-only model in each group); this type of model has been widely studied in econometrics (see Greene, 2000; Muthén, 1983), psychometrics (e.g., Muthén, 1984), and biostatistics (e.g., Ashford & Sowden, 1970; Bock & Gibbons, 1996). This places multivariate SDT within a general statistical framework; a benefit is that the models have been extensively studied in statistics and so a sophisticated methodology is available (the model can be fit using several software packages). For example, the use of the package aML (Lillard & Panis, 2000), which can be used to fit all of the models discussed here, is noted in the appendix; some sample aML programs are available at my website.

Multivariate SDT provides an explicit theory about underlying psychological processes that lead to the observed data. In the examples that follow, the focus is on the parameter estimates, which have a specific interpretation in terms of the theory. It is shown, for example, that the signs of the bivariate correlations of the underlying distributions have both practical and theoretical implications.

2.2. Bivariate SDT and source monitoring: examples

The examples represent recent research on source monitoring where the procedure described above was used, that is, participants gave rating responses for both detection and discrimination. In particular, the data for Experiment 2 of Yonelinas (1999) and Experiments 2 and 3 of Slotnick et al. (2000) are analyzed. In both studies, the sources used were whether a word had been spoken by a woman or a man during the study period. For the test, observers gave a rating response for detection, indicating how confident they were that a word was old or new, and a rating response for discrimination, indicating how confident they were that the source was presented by A or B (i.e., in a female or male voice, respectively). For Yonelinas' study, 1–6 rating responses were used for detection and discrimination, whereas for the studies of Slotnick et al., 1–7 rating responses were used. In both studies, for the discrimination response, higher numbers indicated greater confidence that the source was a male voice (note that the responses for Experiment 2 of Slotnick et al. were reverse coded to be consistent with the others). In the test phase for both studies, items were presented visually.

2.2.1. Goodness-of-fit

Table 1 shows, for all three experiments, likelihood ratio (LR) goodness-of-fit statistics for a fit of the bivariate normal SDT model using FIML estimation in aML; the appendix provides details about calculations of the fit statistics and the degrees of freedom. A goodness-of-fit statistic close in value to the *df* indicates that the null hypothesis that the model fits is not rejected, and so the significant statistics shown in the table indicate that the model does not fit (exactly). It should be recognized, however, that the tests have high power because the sample sizes are quite large. A recognition of this, along with a desire to assess approximate fit in lieu of exact fit, has led to the development of alternative fit indices in statistics. For example, the root mean square error of approximation (RMSEA; Steiger, 1990) is a measure of approximate fit (see the appendix). Of course, what one considers to be approximate fit is subjective, but based on experience with empirical examples, Browne and Cudeck (1993) suggested values of 0.05 or less of RMSEA as indicating close fit, values from 0.05 to 0.08 as indicating acceptable fit, and values greater than 0.10 as indicating poor fit (also see Hu & Bentler, 1999); see the appendix for the formula used to calculate the RMSEA and further comments. The values of RMSEA shown in Table 1 suggest poor fits for the data of Yonelinas (1999) and Slotnick et al. (2000).

Note that the fit statistics are not trustworthy in situations where there are many response patterns (i.e., cells in the multiway table) with expected frequencies less than 5 (see Agresti (2002); having both small and large expected frequencies creates problems as well). This was also noted, in the context of multivariate SDT, by Wickens (1992). Here I note that this problem was present for the data of Slotnick et al. (2000), in that many cells had small expected frequencies (inspection of their published tables shows why, in that the tables are very sparse, with many cells with small or no counts), and was less of a problem for the experiment of Yonelinas (1999).

To summarize, the results suggest that, with respect to applying a bivariate normal SDT model to source monitoring data, there is clearly room for improvement. The marginal receiver operating characteristic (ROC) curves presented by Yonelinas (1999) and Slotnick et al. (2000) suggest a source of some of the lack of fit, in that, for discrimination, the curves on inverse normal coordinates (i.e., z-ROC curves) deviate from linearity, and in particular they are slightly curved (which suggests that some of the lack of fit might be due to violations of distributional assumptions). The curvature could arise for any one of a number of reasons; however, one possibility is suggested by the observation that the curvature is consistent with a mixture extension of SDT for source discrimination (DeCarlo, 2000, in press; also see DeCarlo, 2002), and so it might be worthwhile to develop a multivariate extension of the univariate SDT mixture model. This type of extension is within the realm of SEM, but is outside the scope of the present article; here I simply note that it suggests an interesting direction for future research. Another possibility would be to relax the assumption of decision separability, as in general recognition theory (GRT; Ashby & Townsend, 1986), although this takes the model outside the realm of those developed in SEM (and there are some complexities associated with fitting the model). In future research, a comparison of different types of generalizations of multivariate SDT models, using information criteria for example (Burnham & Anderson, 2002), would be informative.

Table 1
Goodness-of-fit statistics for the bivariate normal SDT model for the data of Yonelinas (1999) and Slotnick et al. (2000)

Study	*LR*	*df*	*p*	*RMSEA*
Yonelinas (Expt. 2, $N = 5760$)	2415.41	84	<0.01	0.120
Slotnick (Expt. 3, $N = 5758$)	3755.09	121	<0.01	0.125
Slotnick (Expt. 2, $N = 2584$)	1579.71	121	<0.01	0.118

Note: Results from aML using FIML. *LR* is the likelihood ratio goodness-of-fit statistic; and *RMSEA* the root mean square error of approximation.

2.2.2. Parameter estimates

Table 2 presents parameter estimates for fits of the bivariate SDT model with ordinal responses to the data of Yonelinas (1999) and Slotnick et al. (2000). The table shows FIML estimates obtained using aML (Lillard & Panis, 2000); the estimates of the standard errors are asymptotic and are referred to as BHHH standard errors in the aML manual (which provides references); aML also offers other estimators of the standard errors (that might be useful for smaller sample sizes). Note that the model as parameterized in aML gives direct estimates of all the parameters shown in Table 2, with the exception that the estimate of d_{AB} is computed using the estimates of $\Psi_{2\mathrm{B}} - \Psi_{2\mathrm{A}}$. Approximate standard errors for d_{AB} were computed by noting that, for

Table 2
Parameter estimates for the bivariate normal SDT model for the data of Yonelinas (1999) and Slotnick et al. (2000)

	d_A	d_B	σ_{1A}	σ_{1B}	d_{AB}	σ_{2A}	σ_{2B}	ρ_N	ρ_A	ρ_B
Yonelinas (Expt. 2)	1.38	1.31	1.47	1.45	1.46	1.85	1.78	−0.02	−0.44	0.48
	(0.06)	(0.06)	(0.06)	(0.06)	(0.08)	(0.07)	(0.06)	(0.02)	(0.03)	(0.03)
Slotnick (Expt. 3)	1.92	1.94	1.42	1.41	2.82	2.80	2.83	−0.06	−0.50	0.51
	(0.07)	(0.07)	(0.06)	(0.06)	(0.15)	(0.10)	(0.11)	(0.04)	(0.04)	(0.04)
Slotnick (Expt. 2)	2.74	2.84	1.90	2.02	4.71	3.32	3.46	0.01	−0.50	0.60
	(0.16)	(0.17)	(0.15)	(0.16)	(0.25)	(0.19)	(0.20)	(0.06)	(0.06)	(0.06)

Note: the table shows FIML estimates obtained using aML. Source A is a female voice, source B is a male voice. The terms in parentheses are the estimated standard errors. The estimate of d_{AB} was obtained as $\Psi_{AN} - \Psi_{BN}$ and the standard error was computed as noted in the text.

random variables X and Y, $V(X-Y) = V(X) + V(Y) - 2\text{Cov}(X, Y)$.

Table 2 shows an impressive consistency of results across the three studies. For example, for all three experiments, the detection of words did not differ across female (A) or male (B) voices, that is, the estimates of the detection parameters d_A and d_B are close in magnitude (and within about a standard error of each other). Second, in all three experiments, the estimates of the standard deviations σ_{1A} and σ_{1B} (for the familiarity dimension) are close in value across A and B and are greater than unity. The finding that σ_{1A} and σ_{1B} are larger than σ_{1N} (which is 1) means that the marginal ROC curves for detection have slopes that are less than unity on inverse normal coordinates, as is usually found. Similarly, the estimates of the standard deviations σ_{2A} and σ_{2B} on the feature dimension are close in value across A and B and are greater than unity; also note that σ_{2A} and σ_{2B} are larger than σ_{1A} and σ_{1B}. Thus, the results for all three experiments suggest that $d_A = d_B$, $\sigma_{1A} = \sigma_{1B}$, and $\sigma_{2A} = \sigma_{2B}$, which suggests that neither detection nor discrimination differed across the two sources (i.e., a female or male voice). In addition, the finding that σ_{2A} and σ_{2B} were larger than σ_{1A} and σ_{1B} suggests that the representations had larger variability on the dimension used for discrimination than on the dimension used for detection.

It should also be noted that Slotnick et al. (2000) used a longer study list in Experiment 3 in order to "reduce the level of performance" (p.1512) and Table 2 shows that the estimates of the detection parameters d_A and d_B and the discrimination parameter d_{AB} are in fact smaller for the longer study list of Experiment 3, as compared to Experiment 2. This provides important experimental evidence in favor of the model, in that it shows that the SDT parameters behave as expected in response to an increase in length of the study list. Note that, although the fits of the multivariate SDT model were poor, the parameter estimates are still informative.

With respect to the bivariate correlations, Table 2 shows that the estimates of ρ_N for the new item distribution are close to zero for all three experiments and are not significant in any case (as can be seen by the size of the standard error), whereas the estimates of ρ_A and ρ_B are large, significant, and opposite in sign. The finding of non-zero values for the estimates of ρ_A and ρ_B means that the detection and discrimination responses were correlated when an old item (A or B) was presented, but not when a new item was presented. The opposite signs arise because high confidence on the discrimination scale involves using response categories that are towards opposite extremes (i.e., 1 = sure A and 6 = sure B). As a result, on a trial where a participant is sure that the item is old, for example, high confidence with respect to discrimination is indicated by high categories (e.g., 5, 6) for B items (and so ρ_B is positive) and low categories (1, 2) for A items (and so ρ_A is negative). It is also interesting to note that, across all three experiments, the absolute magnitude of ρ_A is close to that of ρ_B (and the estimates are well within 2 standard errors).

The finding of non-zero and opposite-signed bivariate correlations for the A and B distributions is a new and interesting result; it also has important implications for situations that involve response selection, as discussed below. It suggests that, when the level of familiarity of an old item was higher, the level of information about features was also higher, perhaps because the item was overall processed at a deeper level. Thus, the results show that information across the two dimensions was correlated. Note that the finding that ρ_N was close to zero in all three studies is consistent with this interpretation, in that a high or low level of familiarity for a new item should not be associated with a higher or lower level of feature information, because new items do not provide information about the source (there is no source associated with new items).

In summary, the results show that the bivariate SDT model of Eq. (3) concisely summarizes major aspects of the data from several source monitoring studies. The results are shown to be consistent across several experiments and the analysis provides new information. Estimates of the parameters of the multivariate SDT model are obtained, as well as estimates of the standard errors; the standard errors are important but have tended to be neglected in prior research (issues of model identification are also important; the use of eigenvalues of the estimated information matrix to help assess identification is noted in the appendix). As discussed above, the parameter estimates have interpretations in terms of detection, discrimination, and response criteria, and the bivariate correlations have implications about underlying processes. Applications of multivariate SDT in future research on source monitoring should be informative.

2.3. Bivariate SDT with rating responses and selection

As noted above, some studies have used a conditional source monitoring procedure, in that a discrimination response was made only if the detection decision was "old". This section shows how to apply multivariate SDT to situations where a conditional source monitoring procedure is used. The most general case where rating responses are given for both detection and discrimination is considered; some sub-models are discussed later in the article. For example, participants might first give a 1–4 detection response and then a 1–4 discrimination response only if the detection response was 3 or 4, as in the example analyzed below.

The bivariate normal SDT model with selection introduced here is an example of what is known more generally in econometrics and statistics as an ordinal-response bivariate probit model with sample selection (e.g., see Greene, 2000; Maddala, 1983). The model presented here extends the usual model in two basic ways: a rating response is used for the selection model (i.e., the detection component) and the model is a multiple group extension of the usual model.

Let $k_1 = (1, \ldots, s, \ldots, K_1)$ indicate K_1 ordered response categories for detection, and $k_2 = (1, \ldots, K_2)$ indicate K_2 ordered response categories for discrimination. In the conditional procedure, a discrimination response is given if and only if $k_1 > s$, and so s is the detection category above which a discrimination response is observed. In this case, the decision rule is

$$Y_{jg} = k_j \quad \text{if } c_{j,k-1} < y^*_{jg} \leqslant c_{jk} \text{ for } j = 1, 2,$$

$$Y_{2g} = k_2 \quad \text{if and only if } Y_{1g} > s \tag{4}$$

for $g = \text{N, A, B}$, and as before, $c_{j0} = -\infty$, $c_{jK} = \infty$, and $c_{j1} < c_{j2} < \cdots < c_{j,K-1}$. The structural model is again Eq. (1) and, with Eq. (4), the resulting model can be written as

$$p(Y_{1g} \leqslant k_1) = \Phi[(c_{1k} - \Psi_{1g})/\sigma_{1g}],$$

$$\begin{aligned} p(&Y_{1g} \geqslant k_1, Y_{2g} \leqslant k_2) \\ &= \Phi_2[(-c_{1,k-1} + \Psi_{1g})/\sigma_{1g}, \\ &\quad (c_{2k} - \Psi_{2g})/\sigma_{2g}, -\rho_g)] \end{aligned} \tag{5}$$

for $k_1 = 1, \ldots, s$ for the first equation and $k_1 = s + 1, \ldots, K_1$ and $k_2 = 1, \ldots, K_2$ for the second equation. Eq. (5) shows that the model consists of two basic components, one for trials where only a detection response is observed and one for trials on which both detection and discrimination responses are observed. In this case, the log likelihood function to be maximized consists of two parts, one for trials where only Y_{1g} is observed and the other for trials where both Y_{1g} and Y_{2g} are observed; the likelihood function follows directly from Eq. (5), again taking differences between the bivariate cumulative probabilities in order to get the probabilities for each response pattern. Thus, the model is easily fit by maximizing the two components of the log likelihood function. Details on fitting bivariate probit models with sample selection can be found in econometrics texts (e.g., see Greene, 1998, 2000). The model was fit here using FIML estimation with the software aML; a sample aML program for bivariate SDT with selection is available at the author's website.

2.4. Bivariate SDT with selection: a source monitoring example

The example is an unpublished pilot source monitoring study conducted by the author. The stimuli consisted of 50 pictures selected from those given by Snodgrass and Vanderwart (1980). Of the 50 pictures, 25 were presented as pictures whereas the other 25 were presented as words. Thus, the two sources were whether the item had been presented as a picture or as a word. For the study period, each picture or word was presented for about 3 s. For the test, participants were tested with either all words or all pictures; in particular, 8 participants were shown 100 words, of which 25 had been shown as words, 25 had been shown as pictures, and 50 were new; whereas 8 other participants were shown 100 pictures, of which 25 had been shown as pictures, 25 had been shown as words, and 50 were new. Although it would be of interest to analyze the data for the two testing conditions separately (i.e., whether participants were tested with pictures or words), the small sample size precludes this and the data pooled across both conditions are analyzed here. Participants gave a 1–6 rating response for detection, followed by a 1–6 rating response for discrimination only if the detection response was greater than 3. For the analysis presented here, categories 2 and 3 (somewhat sure, slightly sure) were combined, as were categories 3 and 4 (because of small counts), resulting in a 4 category scale. Specifically, for detection, the category labels were 1 = sure new, 2 = somewhat or slightly sure new, 3 = somewhat or slightly sure old, and 4 = sure old, whereas for discrimination the labels were 1 = sure word, 2 = somewhat or slightly sure word, 3 = somewhat or slightly sure picture, 4 = sure picture. Note that, in this

Table 3
Parameter estimates for the bivariate normal SDT model with selection, conditional source monitoring data

d_{A}	d_{B}	$\sigma_{1\mathrm{A}}$	$\sigma_{1\mathrm{B}}$	d_{AB}	$\sigma_{2\mathrm{A}}$	$\sigma_{2\mathrm{B}}$	ρ_{N}	ρ_{A}	ρ_{B}
2.16	5.27	1.99	2.06	5.40	2.58	2.42	0.27	−0.90	0.85
(0.16)	(0.79)	(0.20)	(0.46)	(1.17)	(0.62)	(0.74)	(0.37)	(0.04)	(0.10)

Note: the table shows FIML estimates obtained using aML. $N = 1600$. Source A is word, source B is picture. The terms in parentheses are the estimated standard errors. The estimate of d_{AB} was obtained as $\Psi_{\mathrm{AN}} - \Psi_{\mathrm{BN}}$ and the standard error was computed as noted in the text.

case, a discrimination response was observed only if the detection response was 3 or 4, and so $s = 2$ in Eqs. (4) and (5).

An LR goodness-of-fit statistic was computed by substituting the parameter estimates in Eq. (5) and computing the expected frequencies for each response pattern, as noted in the appendix, which gave LR $= 73.52 (N = 1600, df = 10, p < 0.01)$, and so the null hypothesis of exact fit is rejected. The value of RMSEA is 0.109, which suggests a poor to mediocre fit, at best. As before, some of the lack of fit appears to be due to curvature in the marginal ROC curves (which could possibly be handled by a mixture generalization of multivariate SDT).

Table 3 presents the parameter estimates, obtained using FIML in aML. The table shows that the estimate of the detection parameter for pictures (d_{B}, the estimate is 5.27) is considerably larger than that for words (2.16), which suggests a picture superiority effect (see Johnson et al., 1993). The estimate of the discrimination parameter d_{AB} is 5.4, which indicates good discrimination between the sources. The estimates of the standard deviations $\sigma_{1\mathrm{A}}$ and $\sigma_{1\mathrm{B}}$ for the familiarity dimension are both greater than unity and are about equal in value; similarly, the estimates of the standard deviations $\sigma_{2\mathrm{A}}$ and $\sigma_{2\mathrm{B}}$ for the feature dimension are both greater than unity and are about equal in magnitude. The estimates of $\sigma_{2\mathrm{A}}$ and $\sigma_{2\mathrm{B}}$ are also larger than the estimates of $\sigma_{1\mathrm{A}}$, $\sigma_{1\mathrm{B}}$, as was found for the experiments discussed above, which again suggests a difference between recognition memory and source memory. Thus, the general pattern of results for the standard deviations is very similar to that found for the experiments discussed above.

With respect to the bivariate correlations, the estimate of ρ_{N} is small and positive, and is smaller in magnitude than the estimate of its standard error, and so the null hypothesis that ρ_{N} is zero is not rejected, which is consistent with the results found for the three experiments discussed above. The estimates of ρ_{A} and ρ_{B} are large in magnitude, significant, and opposite in sign; the fact that opposite signs are found across procedures with and without selection is important. Overall, the finding of bivariate correlations that are opposite in sign for A and B and near zero for N are again consistent with the view that information used for detection is correlated with information used for discrimination. Finally, it is interesting to note that the estimates of ρ_{A} and ρ_{B} are larger in magnitude than those obtained for the three experiments discussed above (see Table 2), which might reflect an aspect of using pictures versus words as sources in lieu of words spoken in different voices; this and the other interesting findings noted above merit further research.

In summary, the analysis shows that results for a source monitoring study that used rating responses and a conditional procedure were consistent with those obtained for studies that did not use a conditional procedure, in that the estimates of the bivariate correlations for A and B items were large and opposite in sign (and about equal in absolute magnitude) whereas that for new items was not significantly different than zero. Thus, the bivariate normal SDT model with selection offers an informative summary of the data and again reveals new and interesting results. Although both conditional and unconditional versions of the source monitoring procedure are widely used, the present article is the only one I know of that has compared results across the two procedures; additional research on this is needed.

3. On some implications of multivariate SDT for source monitoring

3.1. Conditional source monitoring and bias

It is important to recognize that, from the perspective of multivariate SDT, there is a potentially serious problem associated with the conditional source monitoring procedure. In particular, if the bivariate correlations of the underlying distributions are not zero, then estimates of the discrimination parameter obtained from an analysis that assumes zero correlations might be biased. The bias occurs because of effects of response selection combined with non-zero correlations. In particular, the problem arises because one does not obtain estimates of unconditional means on the feature dimension (dimension 2) with the conditional source monitoring procedure, but rather one obtains estimates of conditional means, because information about $y^*_{2\mathrm{A}}$ and $y^*_{2\mathrm{B}}$ is available only when the realizations of $y^*_{1\mathrm{A}}$ and $y^*_{1\mathrm{B}}$ are greater than the detection criterion c_1 (i.e., $Y_{2\mathrm{A}}$ and $Y_{2\mathrm{B}}$ are observed only when the detection decision is old). For example, for the simple case where an old/new decision is made for detection, it can be shown that the conditional means for discrimination are affected by selection as follows:

$$E(y^*_{2g}|y^*_{1g} > c_1) = \Psi_{2g} + \rho_g \sigma_{2g} \left\{ \frac{\phi[(c_1 - \Psi_{1g})/\sigma_{1g}]}{1 - \Phi[(c_1 - \Psi_{1g})/\sigma_{1g}]} \right\}, \quad (6)$$

where ϕ is the normal probability density function and the rest of the terms are as defined above. Eq. (6) gives the conditional mean of an incidentally truncated bivariate normal distribution, which was derived many years ago (e.g., see Greene, 2000; Kotz, Balakrishnan, & Johnson, 2000). The important aspect of Eq. (6) is that it shows that the mean of the distribution on dimension 2 will be biased in the direction of the correlation (if $\rho_g \neq 0$), which in turn will affect the estimate of the discrimination parameter. For example, whereas the discrimination parameter from the univariate model should be $d_{AB} = \Psi_{2B} - \Psi_{2A}$ (to simplify notation, it is assumed that the subtraction is done so as to give a positive value for d_{AB}), Eq. (6) shows that in the presence of selection the difference in means will instead be

$$d^b_{AB} = d_{AB} + \rho_B \sigma_{2B} \left\{ \frac{\phi[(c_1 - \Psi_{1B})/\sigma_{1B}]}{1 - \Phi[(c_1 - \Psi_{1B})/\sigma_{1B}]} \right\} - \rho_A \sigma_{2A} \left\{ \frac{\phi[(c_1 - \Psi_{1A})/\sigma_{1A}]}{1 - \Phi[(c_1 - \Psi_{1A})/\sigma_{1A}]} \right\}, \quad (7)$$

where d^b_{AB} is the biased value of d_{AB}. Eq. (7) shows that, unless $\rho_g = 0$, the discrimination parameter will be biased if the selection is ignored. The nature of the bias depends on the signs and magnitudes of ρ_A and ρ_B; for example, if ρ_A is negative and ρ_B is positive, as found here, then d^b_{AB} will be inflated. Note that Eq. (7) only considers the effect of truncation on estimation of the mean of a continuous (observed) variable; there might also be additional problems that arise from using categorical observed responses to estimate the mean of the latent variable y^*_{jg}.

Fig. 2 visually illustrates the problem. The figure shows, for new, A, and B items, contours for bivariate normal distributions. The bivariate correlation for the new item distribution is zero, whereas the A and B distributions have non-zero bivariate correlations that are opposite in sign, as found in the experiments analyzed above. A possible location of the criterion on the familiarity dimension is shown (c_1), as well as a possible location of the criterion on the feature dimension (c_2). A conservative detection criterion is shown (i.e., c_1 is far to the right) to help illustrate the point.

Inspection of Fig. 1 should help to show that, for presentations of an A or B item, discrimination will appear to be better when c_1 is far to the right (as in the figure). This occurs because, for a higher criterion, the

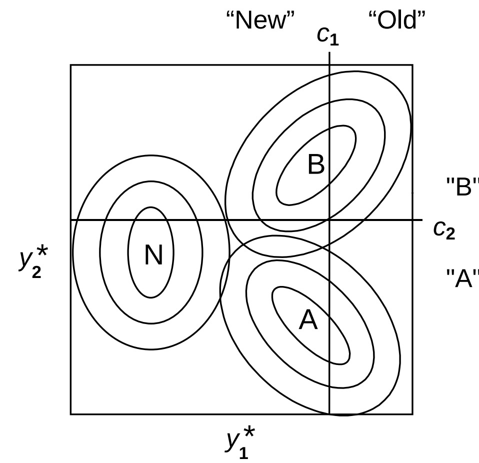

Fig. 2. A figure used to illustrate the effects of response selection. Contours for new, A, and B items for a bivariate SDT model with a yes/no detection decision and an A/B discrimination decision. The familiarity dimension is y^*_1, the feature dimension is y^*_2; c_1 and c_2 are response criteria on dimensions 1 and 2.

portions of the A and B distributions that are sampled tend to be farther apart on the feature dimension (because conditional means are being estimated), and so it will appear as if the A and B distributions are farther apart. Note that if the bivariate correlations for the A and B distributions were zero, then the discrimination estimate would not be biased by selection, as shown by Eq. (7) (the bias might also be minimal if the correlations have the same sign, because the terms in Eq. (7) might largely cancel out, depending on the standard deviations). However, as shown by the analyses presented above, this was not the case for the source monitoring studies examined here, in that the correlations clearly differed from zero and, more importantly, were opposite in sign. As a result, and as shown above by Eq. (7), if one ignores the effects of selection by estimating discrimination in a univariate manner (e.g., by using marginal frequencies or marginal ROC curves), as is commonly done, then the estimates might be biased upwards, which would result in discrimination being over-estimated.

To obtain an idea as to the possible magnitude of the bias, Eq. (7) was used with parameter values that were similar to the estimates obtained in Yonelinas' study (see Table 2). Specifically, the values used were $\rho_A = -0.45$, $\rho_B = 0.45$, $d_{AB} = 1.5$, $d_A = d_B = 1.35$, $\sigma_{1A} = \sigma_{1B} = 1.45$, and $\sigma_{2A} = \sigma_{2B} = 1.8$. Using these values in Eq. (7), c_1 was varied to obtain an idea as to how different locations of the criterion might affect the magnitude of the bias. For example, for $c_1 = -1$, d^b_{AB} was about 1.7, reflecting a small bias of 0.2 (i.e., $1.7 - 1.5$), whereas for $c_1 = 1$, d^b_{AB} was over 2.5, which reflects a quite large bias of 1.0 (note that, for constant d_{AB}, the bias is larger when the criterion is further from the new item distribution, as can be seen in Fig. 1). This shows that the effects of ignoring response selection and simply performing a univariate analysis (which assumes zero bivariate correlations), as is commonly done, can be large enough to be of some concern.

An important implication of Eq. (7) is that it shows that comparisons of discrimination across different groups or conditions, for example, can be distorted by bias if a univariate analysis is used. This has implications for experiments where the conditional source monitoring procedure was used. For example, the conditional procedure was used by Mather et al. (1999) to compare elderly participants to young participants in different conditions (self-focus versus other-focus), where different participants were in different conditions. The procedure was conditional in that participants gave a rating discrimination response only if their (old/new) detection decision was old. In this case, comparisons across the groups based on univariate analyses, such as estimates of the discrimination parameter obtained from the marginal frequencies or a comparison of marginal ROC curves (e.g., see Qin, Raye, Johnson, & Mitchell, 2001) could be misleading if the criterion used for detection differed across the groups, in that what appears to be a difference in discrimination could result from different amounts of bias in the estimates of discrimination. Thus, it is important when making comparisons across groups or conditions based on a univariate analysis to check if the criterion on the detection dimension differs substantially across the groups. If it does not, then although the estimates of the discrimination parameter are still biased, they might not be biased to a different degree across the groups, and so it is possible that comparisons of d_{AB} across groups or conditions are valid (e.g., if the bivariate correlations are similar across groups). If, on the other hand, the criterion locations differ across the groups (or if the bivariate correlations differ substantially), then bias might distort the conclusions.

Of course, if one fits the multivariate model given by Eq. (5) (and allows for non-zero bivariate correlations), then the estimate of d_{AB} is not biased. Note, however, that for the version of the procedure used by Mather et al. (1999) and Hilford et al. (2002), Eq. (5) is not identified. In these studies, participants first decided if an item was old or new and then, only if the decision was old, made a rating response as to how sure they were that the item was presented by source A or source B (the goal was to obtain ROC curves for discrimination). In this case, although a necessary (but not sufficient) condition for identification is satisfied (i.e., there are fewer parameters than observations), the full bivariate SDT model with selection (Eq. (5)) is not identified, because only a binary response was used for detection. The basic problem is that there is no covariation between the detection response and the discrimination response, since the detection response is a constant (i.e., "old") whenever the discrimination response is observed, and so bivariate correlations cannot be estimated. Note that this problem does not arise if (a) a rating response is used for detection and (b) the discrimination response is observed for at least two different values of the rating response, as in the experiment discussed above.

In summary, for the conditional source monitoring procedure, the estimate of the discrimination parameter from a univariate analysis might be biased if the bivariate correlations are not zero, because of the effects of response selection. This raises the possibility that, for a univariate analysis (e.g., comparing marginal discrimination ROC curves across conditions), one cannot be sure that a participant's discrimination differed across the conditions, for example, or if a difference arose because of differential bias across the conditions. This is a potentially serious limitation of using a univariate analysis (or a standard ROC analysis) with data from a conditional source monitoring procedure. Note that the problem of bias does not arise if a univariate analysis is used with data from an unconditional procedure, because there is then no selection. This is an argument in favor of the unconditional procedure over the conditional procedure.

3.2. Dichotomous responses and identifiability

This section considers situations where only dichotomous responses (e.g., yes or no, A or B) are used, as in many source monitoring studies. Once again, there are two basic versions of the procedure, corresponding to the two versions discussed above. In one version, participants first give an old or new response, and then give an A or B response, regardless of their first response. In the second procedure, participants only give an A or B response when the first response is old.

For the first situation, which does not involve response selection, the appropriate model is simply Eq. (3) with dichotomous responses. It is important to recognize, however, that the full model is not identified because a necessary condition for identification, which is that the number of parameters is less than or equal to the number of observations, is not satisfied. In particular, there are a total of 9 observations: there are 4 possible response patterns for each item (i.e., new and A, new and B, old and A, old and B) and, because the number of presentations of each item is fixed by design, the frequencies for only 3 of the 4 response patterns are free to vary. Thus, for 3 items (new, A, B), there are a total of $3 \times 3 = 9$ free frequencies. The multivariate SDT model, however, has 13 parameters (i.e., two response criteria, one on each dimension, and d_A, d_B, Ψ_{2A}, Ψ_{2B}, σ_{1A}, σ_{1B}, σ_{2A}, σ_{2B}, ρ_N, ρ_A, ρ_B). Thus, the full model is not identified, because there are 13 parameters but only 9 observations. Note that a sub-model where the standard deviations are set to unity is identified (the model then has 9 parameters and so is exactly identified), and so one can still estimate the bivariate correlations in this case, but only with the assumption that the standard deviations are unity (or possibly by imposing another parameter restriction).

The situation is even worse when dichotomous responses are used with the conditional source monitoring procedure. The appropriate model in this case is Eq. (5), which has 13 parameters. With respect to the number of observations, there are now only 3 possible response patterns for each item (new, old and A, old and B). The total number of presentations for each item is again fixed by design, and so the frequencies for only 2 of the 3 patterns are free to vary. For 3 items, this gives a total of $3 \times 2 = 6$ free frequencies. Thus, the model is not identified, in that there are 13 parameters but only 6 observations. In this case, a sub-model with standard deviations of unity and zero bivariate correlations is identified; the sub-model is equivalent to simply fitting separate equal-variance univariate SDT models to the marginal frequencies. Of course, if the bivariate correlations are not zero, then the estimate of the discrimination parameter might be biased, as discussed above.

Note that the conditional procedure with dichotomous responses is similar to a univariate version of the source monitoring procedure where participants simply give one response on each trial, namely "new", "A", or "B"; Macmillan and Creelman (1991) made a similar observation with respect to a version of simultaneous detection and identification (see p.239). Thus, an important implication of multivariate SDT for the univariate-version of the source monitoring procedure is that if one simply uses the A and B responses to estimate discrimination (e.g., from the bivariate frequency table for A and B responses), then the estimate might be biased because of effects of selection.

Finally, it should be emphasized that the problems with identifiability noted above that arise when dichotomous responses are used are easily avoided by using rating responses. The use of rating responses allows one to fit the full models of Eq. (3) or (5), and therefore one does not have to make (possibly false) assumptions about the bivariate correlations (i.e., that they are zero) or the standard deviations (i.e., that they are unity). This is an important take-home message for researchers interested in source monitoring.

4. Conclusions

Source monitoring studies, and similar designs, are widely used in psychology, and there is clearly a need for theoretically motivated models to help organize, summarize, and interpret the data. The present article shows that multivariate SDT provides a useful framework for source monitoring. In particular, applications of rating response bivariate SDT models with and without selection to recent studies are shown to provide interesting and informative results. It is emphasized that it is important to pay attention to the particular design that is used and the type of responses that are given. The application of bivariate SDT to source monitoring also helps to clarify limitations of univariate analyses and the conditions under which they can validly be used.

The approach to source monitoring through multivariate SDT also suggests directions for future research. For example, it would be interesting to experimentally manipulate, in a rating response experiment, item and source similarity, as in a study of Bayen, Murnane, and Erdfelder (1996); (where a univariate procedure was used). In addition to seeing if the detection and discrimination parameter estimates behaved appropriately, this would allow one to see if the bivariate correlations of the A and B distributions are affected by experimental manipulations of item or source similarity (note that the results for Experiments 2 and 3 of Slotnick et al. shown above suggest that detection and discrimination can be increased without affecting the bivariate correlations). It would also be of interest to conduct additional studies that compare results across procedures with and without selection. Overall, the approach to source monitoring via multivariate SDT is informative with respect to both design of the study and analysis of the data. The ability to fit the models with widely available software should also help to encourage researchers to use them.

Appendix

Some notes on using the software aML (Lillard & Panis, 2000) and on the computation of the goodness-of-fit statistics and their *df* are given here. aML can be used to fit both the bivariate SDT model given by Eq. (3) and the bivariate SDT model with selection given by Eq. (5), using full information maximum likelihood (FIML) estimation. Sample aML programs for some of the examples discussed here are available at the author's website.

The software aML (Lillard & Panis, 2000) is a package for fitting multilevel multiprocess models. It can be used to fit multivariate ordinal-response models, using FIML, for up to three indicators (for more than three indicators it uses marginal maximum likelihood). There are actually several statistics (e.g., Stata; StataCorp., 1999) and econometrics (e.g., LIMDEP; Greene, 1998) packages available that allow one to fit bivariate probit models with and without selection, however some programming is required to extend the models to multiple groups. This is quite easily accomplished with aML, since it is designed specifically for multilevel multiprocess models.

When using aML, the use of good starting values is important; a good strategy is to start with a simple version of the model and to then build up to more complex models. For example, one can first fit univariate probit models separately to the marginal detection and discrimination responses in order to obtain starting values for the bivariate model. The bivariate model can then be fit with unit variances and zero bivariate correlations; next, the variances can be freed (except for the reference distribution) and the estimates obtained can be used as starting values for a fit of the full model with free bivariate correlations.

It is also important to check information about identification, which can be done using the eigenvalues of the estimated information matrix; the five smallest eigenvalues are given in the aML output. Zero or near zero values for the eigenvalues indicate that the model is not identified. This was checked for all of the models presented here and was not a problem in any case.

Goodness-of-fit statistics are not reported by aML, but they can be calculated using the parameter estimates given in the output. In particular, substituting maximum likelihood estimates of c_{jk}, Ψ_{jg}, σ_{jg}, and ρ_g into the terms on the right-hand side of Eq. (3) or (5) gives expected cumulative probabilities; the probabilities for each response pattern and item can then be obtained by subtracting the appropriate cumulative probabilities. Multiplying the expected probabilities by the sample size for each item gives the expected frequencies. The expected and observed frequencies for each response pattern and item can then be used to compute the LR (or chi-square) goodness-of-fit statistic in the usual manner. For example, the LR statistic is computed as

$$LR\ X^2 = 2\sum O \ln(O/\hat{E}),$$

where the summation is over all the response patterns, ln is the natural logarithm, O represents the observed frequencies, and E-hat are the estimated expected frequencies.

With respect to calculating the df for goodness-of-fit statistics with FIML, the number of response patterns determine the number of observations. For example, for the experiment of Yonelinas (1999) with two 1–6 rating responses, there are 35 free frequencies in the 6×6 joint table of responses for each item (one frequency is not free because the table total for each item is fixed by design). For 3 items, this gives a total of 105 free joint frequencies (observations), and so the df are obtained by subtracting the number of model parameters from 105. For the bivariate SDT model, there are 21 parameters (5 criteria on each of 2 dimensions, 4 free means, for A and B, 4 standard deviations, again for A and B, and 3 bivariate correlations) which gives $105 - 21 = 84$ df, as shown in Table 1.

For the example involving response-selection with two 1–4 rating responses, there are 2 response patterns where only Y_{1g} is observed (i.e., responses of 1 or 2 for detection) and 8 response patterns where both Y_{1g} and Y_{2g} are observed (i.e., responses of 3 or 4 for detection combined with responses from 1–4 for discrimination), which gives a total of 10 response patterns for each item. Frequencies for 9 of the 10 patterns are free (the number of item presentations is fixed by design), and so for 3 items there are $3 \times 9 = 27$ observations. For the bivariate probit model with sample selection, there are 17 parameters (3 criteria on 2 dimensions, 4 means, 4 standard deviations, and 3 bivariate correlations), and so the df are $27 - 17 = 10$, as reported in the text.

As noted in the text, bivariate normal SDT is basically a multiple group bivariate probit model, with the groups being the different items. Thus, a multiple group version of RMSEA was used for the results reported in the text (see Steiger, 1998),

$$RMSEA = \sqrt{\max\left(\frac{X^2}{N\,df} - \frac{1}{N}, 0\right)} \times \sqrt{g},$$

where X^2 is the value of the chi-square goodness-of-fit statistic, N is the total sample size, df are the degrees of freedom for the fitted model, and g is the number of groups, which is 3 for source monitoring with 2 sources. Note that RMSEA was developed in the context of models with continuous indicators; research on its use with models with categorical indicators is needed. One can also obtain a confidence interval for RMSEA, at least for continuous data (this also needs to be investigated for categorical data).

References

Agresti, A. (2002). *Categorical data analysis* (2nd ed.), Hoboken, NJ: John Wiley and Sons.

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

Ashby, F. G., & Townsend, J. T. (1986). Varieties of perceptual independence. *Psychological Review*, *93*, 154–179.

Ashford, J. R., & Sowden, R. R. (1970). Multivariate probit analysis. *Biometrics*, *26*, 535–546.

Banks, W. P. (2000). Recognition and source memory as multivariate decision processes. *Psychological Science*, *11*, 271–277.

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

Batchelder, W. H., & Riefer, D. M. (1999). Theoretical and empirical review of multinomial process tree modeling. *Psychonomic Bulletin & Review*, *6*, 57–86.

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

Bock, R. D., & Gibbons, R. (1996). High dimensional multivariate probit analysis. *Biometrics*, *52*, 1183–1194.

Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley.

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

Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach* (2nd ed.). New York: Springer.

DeCarlo, L. T. (2000). *An extension of signal detection theory via finite mixtures of generalized linear models*. Paper presented at the 2000 Meeting of the Society for Mathematical Psychology, Kingston, Ontario. Available at http://www.columbia.edu/~ld208.

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

DeCarlo, L. T. (in press). An application of signal detection theory with finite mixture distributions to source discrimination. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.

Greene, W. H. (1998). *Limdep version 7.0 user's manual* (revised edition). New York: Econometric Software.

Greene, W. H. (2000). *Econometric analysis* (4th ed.). Upper Saddle River, NJ: Prentice-Hall.

Hilford, A., Glanzer, M., Kim, K., & DeCarlo, L. T. (2002). Theories of source memory: ROC analysis. *Journal of Experimental Psychology: General*, *131*, 494–570.

Hu, L.-T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*, 1–55.

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.

Kotz, S., Balakrishnan, N., & Johnson, N.L. (2000). *Continuous multivariate distributions*, Vol. 1. *Models and applications* (2nd ed.). New York: Wiley.

Lillard, L. A., & Panis, C. W. A. (2000). *aML user's guide and reference manual*. Los Angeles, CA: EconWare Inc.

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

Maddala, G. S. (1983). *Limited-dependent and qualitative variables in econometrics*. New York: Cambridge University Press.

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

Muthén, B. O. (1983). Latent variable structural equation modeling with categorical data. *Journal of Econometrics*, *22*, 43–65.

Muthén, B. O. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika*, *49*, 115–132.

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

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

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory*, *6*, 174–215.

StataCorp. (1999). *Stata statistical software*: Release 6.0. College Station, TX: Stata Corporation.

Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research*, *25*, 173–180.

Steiger, J. H. (1998). A note on multiple sample extensions of the RMSEA fit index. *Structural Equation Modeling*, *5*, 411–419.

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

Wickens, T. D. (1992). Maximum-likelihood estimation of a multivariate Gaussian rating model with excluded data. *Journal of Mathematical Psychology*, *36*, 213–234.

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