Toward a complete decision model of item and source recognition: A discrete-state approach

KARL CHRISTOPH KLAUER
Albert-Ludwigs-Universität Freiburg, Freiburg, Germany

AND

DAVID KELLEN
University of Lisbon, Lisbon, Portugal

In source-monitoring experiments, participants study items from two sources (A and B). At test, they are presented Source A items, Source B items, and new items. They are asked to decide whether a test item is old or new (item memory) and whether it is a Source A or a Source B item (source memory). Hautus, Macmillan, and Rotello (2008) developed models, couched in a bivariate signal detection framework, that account for item and source memory across several data sets collected in a confidence-rating response format. The present article enlarges the set of candidate models with a discrete-state model. The model is a straightforward extension of Bayen, Murnane, and Erdfelder's (1996) multinomial model of source discrimination to confidence ratings. On the basis of the evaluation criteria adopted by Hautus et al., it provides a better account of the data than do Hautus et al.'s models.

Being able to recognize previously acquired information is one of the best-studied abilities in the memory literature, an ability often referred to as *item memory* (for a review, see Malmberg, 2008). But being able to recognize something that was previously encountered represents only a part of a more elaborated faculty. People can often also remember the origin and specific characteristics of previously encountered stimuli, an ability that is often referred to as *source memory* (Johnson, Hashtroudi, & Lindsay, 1993). Experiments on source memory involve the discrimination of one or more characteristics of previously studied items, such as whether a given item was spoken by a female voice (Source A) or by a male voice (Source B). Just as for item memory, source judgments can be collected using a binary responses format (e.g., respond "male" or "female") or by means of a rating scale expressing graded degrees of confidence in a source ascription.

Source memory is usually not studied separately from item memory (for an exception, see DeCarlo, 2003a). Prior to the source judgment, participants typically judge whether the item was previously presented or not. Traditionally, source judgments are requested only for items judged old. The joint assessment of item and source memory has two main advantages. First, it permits a comprehensive assessment of memory performance in different populations, such as young and healthy individuals (e.g., Dodson & Shimamura, 2000; Johnson, Kounios, & Reeder, 1994), elderly individuals (e.g., Spencer & Raz,

1995), and specific clinical populations (e.g., Multhaup & Balota, 1997). Second, it allows one to investigate how the processes underlying both judgments interact—for example, in shaping guessing biases (Bayen, Nakamura, Dupuis, & Yang, 2000; Johnson & Raye, 1981; Meiser, Sattler, & von Hecker, 2007; Riefer, Hu, & Batchelder, 1994)—affecting overall performance.

One approach to studying source memory is the source-monitoring model proposed by Batchelder and colleagues (Batchelder & Riefer, 1990; Batchelder, Riefer, & Hu, 1994; Riefer et al., 1994), a discrete-state model that has successfully been used to distinguish source memory from both item memory and several guessing processes involved (but see Kinchla, 1994). This model has been further developed and extended to deal with different aspects of item and source memory, such as distractor detection (Bayen, Murnane, & Erdfelder, 1996), multiple source dimensions (Klauer, Ehrenberg, & Wegener, 2003; Meiser, 2005; Meiser & Bröder, 2002), and partial memory states (Dodson, Holland, & Shimamura, 1998; Klauer & Wegener, 1998).

Distinguishing Between Source Memory Models by Means of ROC Analyses

More recently, the discrete-state model has been considered less adequate than models based on continuous processes, such as signal detection models (Glanzer, Hilford, & Kim, 2004), or than hybrid models that assume a combination of continuous and threshold processes (Yonelinas,

465

1997, 1999). This preference is based on research in which confidence ratings for source memory were collected and used to compile so-called receiver-operating characteristic (ROC) functions (Glanzer et al., 2004; Hilford, Glanzer, Kim, & DeCarlo, 2002; Qin, Raye, Johnson, & Mitchell, 2001; Slotnick & Dodson, 2005; Slotnick, Klein, Dodson, & Shimamura, 2000; Yonelinas, 1999). The use of ROC functions is familiar from studies of item memory (for reviews, see Wixted, 2007; Yonelinas & Parks, 2007), in which discrete-state models predict linear ROCs (for binary response formats), whereas signal detection models predict curvilinear ones.

According to the hybrid model proposed by Yonelinas (1999), item memory judgments are based on both familiarity processes (defined by a signal detection model) and recollection processes (defined as a threshold process). In this model, source judgments are based on recollection for sources equated in familiarity, leading to linear ROCs. For sources differing in familiarity, familiarity is a diagnostic feature in source discrimination, leading to curvilinear ROCs. In order to test these predictions, Yonelinas (1999) compared source ROCs for sources equated in familiarity with source ROCs for differentially familiar sources. The results were generally consonant with the model predictions, with linear source ROC functions (pooled across old/new confidence ratings) when the familiarity levels of both Sources A and B were equivalent. When familiarity differed between sources, the observed ROC was curvilinear, reflecting the operation of a familiarity process, according to Yonelinas (1999).

The claim of linear source ROCs for equally familiar sources was, however, questioned by Qin et al. (2001), who observed curvilinear ROC functions. Slotnick et al. (2000) also reported curvilinear source ROCs (pooled across old/new confidence ratings), albeit with a lesser degree of curvature than typically observed for item memory ROCs (see also Hilford et al., 2002). Although this flattening is problematic for the Gaussian assumption of SDT (Green & Swets, 1966), these results broadly support a model in which both item and source judgments are based on continuous processes.

One of the issues raised by Slotnick et al. (2000) was that the shape of the source ROCs depended on the previous old/new rating. Slotnick et al. presented so-called refined source ROCs computed conditionally on fixed levels of confidence in the old/new response. Source ROCs for the most confident *old* responses were highly curvilinear, whereas source ROCs based on lower levels of old/new confidence were increasingly flattened (see also the present Figure 3, lower right panel). This direct relationship between curvature and old/new confidence rating was ascribed to "guessing noise" by Slotnick and Dodson (2005), a potential source of distortion of ROCs originally proposed by Ratcliff, McKoon, and Tindall (1994). Through a reanalysis of the data from Yonelinas (1999) and two additional experiments in which memory strength was manipulated, Slotnick and Dodson showed that the amount of noise appeared to increase with decreasing memory strength, as participants become less likely to remember

any diagnostic source information. This interpretation of the flattening of source ROCs as being due to guessing noise allows the signal detection approach to account for the seemingly incongruent ROCs found in the literature.

An alternative yet similar account was previously proposed by DeCarlo (2003a; see also DeCarlo, 2002). In this *mixture distribution* signal detection model, there is an additional signal distribution to describe items that are presented but not attended to during study. Participants have no source information available for items that they did not attend to, and so the distribution of unattended targets contains no systematic source information. In this model, targets can thus enter one of two discrete states: They can have been attended to during study or not. As in a discrete-state model, a mixture parameter governs the probability with which targets enter either state. This mixture account provides a better description of source ROCs than does a pure signal detection model, using an additional mixture parameter that is theoretically driven and psychologically plausible.

One common thread of these studies is that models based only on threshold processes are not considered competitive, given the curvilinear ROCs found in item and source memory. The discussion is, instead, focused on the question of whether a signal detection model augmented by an additional discrete process (recollection or mixture distribution) can reproduce the observed ROCs. The dismissal of discrete-state models on the basis of curvilinear ROCs relies on the idea that threshold models predict linear ROCs not only for binary responses (Macmillan & Creelman, 2005; Murdock, 1974), but also when rating scales are used (e.g., Hilford et al., 2002). Actually, a debate regarding the representation of rating responses in threshold models pointed out long ago that this is a misconception (e.g., Broadbent, 1966; Krantz, 1969; Larkin, 1965; Lee, 1963; Nachmias & Steinman, 1963; Watson & Bourbon, 1965; Wickelgren, 1968).

As several authors have argued before (e.g., Bröder & Schütz, 2009; Erdfelder & Buchner, 1998; Falmagne, 1985; Krantz, 1969; Malmberg, 2002), discrete-state models must specify how mental states (such as a mental state with item memory, but without source memory) map onto the available responses. For binary old/new and source decisions, the issue of response mapping is trivial, as elaborated below, but for graded response formats, a more complex mapping of states on responses is needed. Malmberg (2002) demonstrated that a two-high threshold model can thereby produce curvilinear ROCs, including the asymmetrical functions predicted by signal detection models with unequal variances and by Yonelinas's (1999) dual-process model. In the domain of item memory, Bröder and Schütz recently showed that ROC functions obtained with binary old/new responses and experimental manipulations of old/new response bias have a linear shape, as predicted by a discrete-state model. These authors argued (1) that it may be premature to dismiss discrete-state models on the basis of curvilinear ROCs defined by confidence ratings and (2) that the shape of the ROCs based on confidence ratings may not be diagnostic.

**Toward an Extended Model for
Item and Source Memory**

The signal detection models discussed so far did not attempt to account for both item and source memory ROCs simultaneously. In order to account for both item and source memory and corresponding ROCs, DeCarlo (2003b) proposed a two-dimensional signal detection model (see also Ashby, 1988, 1992; Banks, 2000; Kinchla, 1994; Tanner, 1956; Wickens, 1992). The model postulates three two-dimensional normal distributions: one for new items, one for studied items from Source A, and one for studied items from Source B. One axis in the two-dimensional space represents item memory, the second source memory. Like its one-dimensional counterpart, the two-dimensional signal detection model assumes linear response bounds for each of the memory judgments involved. Despite capturing the overall pattern of the Slotnick et al. (2000) and Yonelinas (1999) data sets, the overall goodness of fit (evaluated, e.g., through the G^2 measure of goodness of fit described below) was poor. The DeCarlo (2003b) model does, however, represent the first attempt to model both types of ROCs simultaneously.

Its limitations were highly informative for Hautus, Macmillan, and Rotello (2008), who implemented a series of modifications to the original model in a stepwise fashion, greatly increasing its ability to account for the three data sets considered—that is, for the data from Yonelinas's (1999) Experiment 2 and for the data from Slotnick et al.'s (2000) Experiments 2 and 3. These data sets have repeatedly been used in developing comprehensive models of item and source memory.

The first modification was the use of likelihood ratio bounds, instead of linear response bounds, for source ratings. This modification (Model 1) resulted in a substantial increase in goodness of fit, and it accounted for the observed flattening of refined source ROCs contingent on level of old/new confidence. Model 1 also accommodated source ROCs that were curved when represented in *z*-coordinates, which, until then, only hybrid models could predict.

In a second step, Hautus et al. (2008) proposed that ratings corresponding to a *new* response are followed by simple source guessing. The implementation of this idea (Model 2) led to another substantial increase in goodness of fit.

The final modification to DeCarlo's (2003b) two-dimensional model was to introduce a noise distribution modeling old items that were not attended to during study. The resulting model (Model 3) further increased goodness of fit and outperformed the other models even when the increased complexity of the model was taken into account by means of model selection indices that penalize a model for its complexity. This final model represents a tremendous improvement in the ability of the signal detection framework to account for both item and source ROCs, and it is the best account in the literature so far.

A related model was, however, recently proposed by Onyper, Zhang, and Howard (2010). As in Hautus et al.'s (2008) Model 3, studied items are represented by a mix-

ture of two distributions. But these are not identified with nonattended and attended targets, respectively; instead, one distribution is said to represent a state of familiarity without source memory, the other one a state of recollection with diagnostic item and source information. There are thus three normal distributions for studied items: one for studied items from Source A, one for studied items from Source B, and a third distribution representing a state of familiarity with lowered mean on the axis representing item memory and without diagnostic source information. Source and item information are assumed to vary independently in each distribution. Response bounds are linear, but response bounds for the source judgment are estimated separately for each level of confidence in the old/new response.

Hautus et al. (2008, p. 892) were motivated by the desire to extend the menu of models from which researchers can choose in modeling given data, in order to increase the chances of finding a model that is capable of describing all the data sets considered. In the present article, we follow this research strategy, adding a discrete-state model to the set of candidate models. We evaluate the performance of this model in the enlarged set of models, using the same statistical criteria and ROC analyses as those adopted by Hautus et al. for the purpose. We also fitted Onyper et al.'s (2010) model.

THE DISCRETE-STATE MODEL

We build on the well-known multinomial model of source monitoring originally proposed by Batchelder and Riefer (1990) in the two-high threshold version developed by Bayen et al. (1996). Consider first the case of dichotomous responses. In a typical source-monitoring experiment, participants acquire information from two sources, Source A and Source B. At test, the participants are presented Source A items, Source B items, and new items in a mixed list. For each item, they are asked to decide whether it is old or new. Given an *old* response, they are to decide whether it stems from Source A or B. In the case of a *new* response, source decisions are typically not required.

It is convenient to present the model in two parts, a stimulus–state mapping and a state–response mapping.

Stimulus–State Mapping

In Bayen et al.'s (1996) model, we can distinguish five mental states, M_1 to M_5 (see Figure 1):

M_1: For an A item, it is remembered that the item is old and from Source A.

M_2: For a B item, it is remembered that the item is old and from Source B.

M_3: For an old item, it is remembered that the item is old, but memory for the source is absent.

M_4: For a new item, it is detected that it is new.

M_5: For an item presented at test, it is not remembered that the item is old, and it is not detected that the item is new.

Stimulus–State Mapping
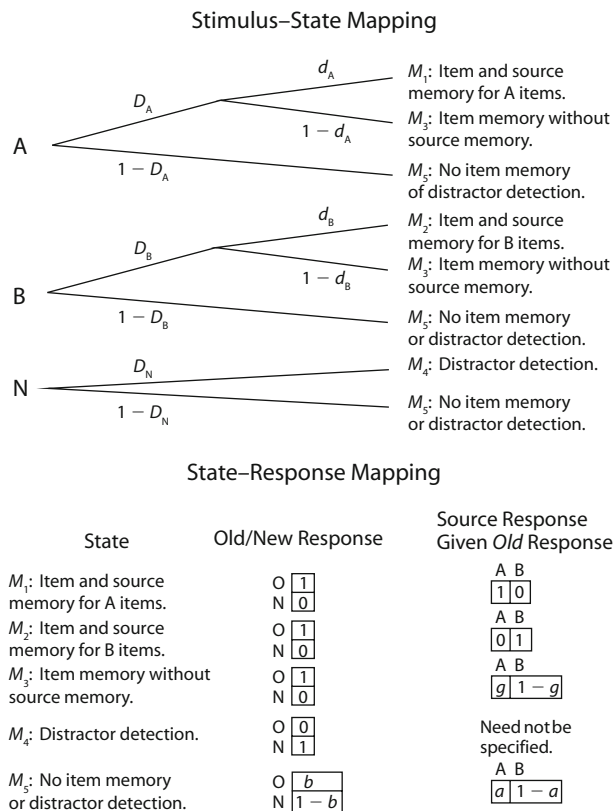

State–Response Mapping

Figure 1. The source-monitoring model with state–response mapping for binary data. The upper part of the figure shows the stimulus–state mapping. A refers to a target from Source A, B to a target from Source B, N is a distractor item. The lower part shows the state–response mapping.

Distractor detection as described in state M_4 is believed to rely on metacognitive inference. For example, seeing a new item, participants may infer that it is new on the basis of, for example, some unusual feature of it, idiosyncratic associations it evokes, or its overall incongruency with the set of targets (Strack & Bless, 1994).

The stimulus–state mapping shown in the upper part of Figure 1 describes the probabilities of entering each of these states given an A item, a B item, or a new item. It uses two item memory parameters D_A and D_B, one distractor detection parameter D_N, and two source memory parameters d_A and d_B. They are defined as follows:

D_A: the probability of remembering that an A item is old;

D_B: the probability of remembering that a B item is old;

D_N: the probability of detecting that a new item is new;

d_A: the probability of remembering the source of an A item, given item memory;

d_B: the probability of remembering the source of a B item, given item memory.

For example, according to Figure 1, the probability of entering state M_1 for an A item is given by $D_A \times d_A$.

State–Response Mapping for Dichotomous Data

The model is completed by adding the state–response mapping that specifies how the mental states map onto the observable responses. The state–response mapping is shown in the lower part of Figure 1. For each mental state, it specifies first the probability with which an *old* or *new* response will be elicited in that state (column labeled "Old/New Response"), followed by the conditional probability of a Source A or B response given an *old* response (column labeled "Source Response Given *Old* Response").

In mental state M_1, there is both item and source memory for a presented A item. Hence, the *old* response has a probability of one, the *new* response a probability of zero. Following the *old* response, the response A is given with a probability of one, and the response B with a probability of zero.

In mental state M_3, there is item memory, leading to *old* responses with a probability of one; but there is no source memory, so Source A is guessed with a probability of g, and Source B with a probability of $1 - g$.

In mental state M_4, it is detected that a new item is new, and the *new* response is given with a probability of one. Because only *new* responses arise in this mental state, the response mapping for the source decision need not be specified (a source decision is typically not required for *new* responses).

In mental state M_5, there is neither item nor source memory. The old/new response is then guessed with a probability of b for an *old* response and $1 - b$ for a *new* response. In the case of an *old* response, a source decision is required, and Source A is guessed with a probability of a and Source B with a probability of $1 - a$. Frequently, source guessing with item memory (state M_3) and source guessing without item memory (state M_5) are described by the same parameter; that is, g is set equal to a.

Note that guessing as reflected in parameters a, b, and g is not necessarily random guessing. Guessing comprises a variety of informed guessing strategies influenced by various sources of information such as base rates, metacognitive inferences, strategies, and so forth, in the absence of a clear memory-based signal as to the appropriate response.

State–Response Mapping for Rating Data

Can the model for dichotomous responses be extended to paradigms in which item and/or source decisions are collected by means of confidence ratings? Consider, for example, an experiment with 6-point rating scales for item recognition and source discrimination, in which participants judge each item twice, once with respect to whether it was old (responses 6, 5, and 4 in order of decreasing confidence) or new (responses 3, 2, and 1 in order of increasing confidence), and once with respect to whether it stemmed from Source A (responses 1, 2, and 3 in order of decreasing confidence) or Source B (responses 4, 5, and 6 in order of increasing confidence). Thus, each response falls into one of the cells of a 6 × 6 matrix, with rows coding the old/new rating and columns the source rating. The data, aggregated across items, can be summarized by three such matrices: one for A items, one for B items, and one

for new items. Cell entries are the frequencies with which the combinations of old/new rating and source rating were observed. The data by Yonelinas (1999, Experiment 2) follow this format.

It is, in fact, straightforward to extend the model to this response format. The stimulus–state mapping is unchanged, because it is independent of response format. In consequence, only the state–response mapping must be extended, which can be done in a natural fashion.

Table 1 shows the appropriate state–response mapping. As for the dichotomous case, the mapping specifies the probabilities of the different old/new responses (column labeled "Old/New Parameters") and the conditional probabilities of the different source-rating responses given one of the possible old/new ratings for each mental state (columns headed "Source Rating Parameters").

In mental state M_1, there is both item memory and source memory for Source A. Because there is item memory, the *new* responses on the old/new scale (i.e., 1, 2, and 3) receive a probability of zero, as in the dichotomous case. There are, however, three *old* responses (6, 5, and 4) that can occur, and the probabilities of using these in state M_1 are modeled by three parameters, $r(6|M_1)$, $r(5|M_1)$, and $r(4|M_1)$, one for each *old* response. The three r parameters have to sum to one, so that there are only two free parameters to be estimated. For reasons elaborated below (i.e., for variations in response style; see the Simplifying the State–

Response Mapping section), it is not reasonable to assume that detect states such as M_1 are invariably mapped onto highest confidence responses (see Onyper et al., 2010, for a similar assumption in the signal detection framework).

In mental state M_1, there is also memory for Source A; hence, the Source B responses on the source-rating scale (i.e., 4, 5, and 6) receive a probability of zero, as in the dichotomous case. Conditional on each of the three *old* responses that can arise in M_1, the parameters s specify the probabilities of the three Source A responses. For example, $s(1|5)$ is the probability of using source-rating category 1 given an old/new rating of 5. There are now three rows of s parameters, because the s parameters are defined conditionally on the actual *old* response. In each row, the three s parameters have to sum to one, so that each row contributes only two free parameters. The source-rating parameters for the rows with *new* responses need not be specified, because *new* responses do not occur in state M_1.

In mental state M_3, there is item memory without source memory. The source-rating parameters in state M_3 have been named g_o, because they correspond to the source-guessing parameter g in the dichotomous case.

In mental state M_4, it is detected that a new item is new. Unlike in the dichotomous case, source ratings are now also available for *new* responses; hence, new parameters for source guessing are required conditional on the *new* responses in state M_4. They were named g_n.

Table 1
State–Response Mapping

Mental State	Old/New Rat.	Old/New Par.	Source Rating Parameters 1	2	3	4	5	6
M_1	6	$r(6|M_1)$	$s(1|6)$	$s(2|6)$	$s(3|6)$	0	0	0
	5	$r(5|M_1)$	$s(1|5)$	$s(2|5)$	$s(3|5)$	0	0	0
	4	$r(4|M_1)$	$s(1|4)$	$s(2|4)$	$s(3|4)$	0	0	0
	3	0						
	2	0			need not be specified			
	1	0						
M_2	6	$r(6|M_2)$	0	0	0	$s(4|6)$	$s(5|6)$	$s(6|6)$
	5	$r(5|M_2)$	0	0	0	$s(4|5)$	$s(5|5)$	$s(6|5)$
	4	$r(4|M_2)$	0	0	0	$s(4|4)$	$s(5|4)$	$s(6|4)$
	3	0						
	2	0			need not be specified			
	1	0						
M_3	6	$r(6|M_3)$	$g_o(1|6)$	$g_o(2|6)$	$g_o(3|6)$	$g_o(4|6)$	$g_o(5|6)$	$g_o(6|6)$
	5	$r(5|M_3)$	$g_o(1|5)$	$g_o(2|5)$	$g_o(3|5)$	$g_o(4|5)$	$g_o(5|5)$	$g_o(6|5)$
	4	$r(4|M_3)$	$g_o(1|4)$	$g_o(2|4)$	$g_o(3|4)$	$g_o(4|4)$	$g_o(5|4)$	$g_o(6|4)$
	3	0						
	2	0			need not be specified			
	1	0						
M_4	6	0						
	5	0			need not be specified			
	4	0						
	3	$r(3|M_4)$	$g_n(1|3)$	$g_n(2|3)$	$g_n(3|3)$	$g_n(4|3)$	$g_n(5|3)$	$g_n(6|3)$
	2	$r(2|M_4)$	$g_n(1|2)$	$g_n(2|2)$	$g_n(3|2)$	$g_n(4|2)$	$g_n(5|2)$	$g_n(6|2)$
	1	$r(1|M_4)$	$g_n(1|1)$	$g_n(2|1)$	$g_n(3|1)$	$g_n(4|1)$	$g_n(5|1)$	$g_n(6|1)$
M_5	6	$b(6)$	$a_o(1|6)$	$a_o(2|6)$	$a_o(3|6)$	$a_o(4|6)$	$a_o(5|6)$	$a_o(6|6)$
	5	$b(5)$	$a_o(1|5)$	$a_o(2|5)$	$a_o(3|5)$	$a_o(4|5)$	$a_o(5|5)$	$a_o(6|5)$
	4	$b(4)$	$a_o(1|4)$	$a_o(2|4)$	$a_o(3|4)$	$a_o(4|4)$	$a_o(5|4)$	$a_o(6|4)$
	3	$b(3)$	$a_n(1|3)$	$a_n(2|3)$	$a_n(3|3)$	$a_n(4|3)$	$a_n(5|3)$	$a_n(6|3)$
	2	$b(2)$	$a_n(1|2)$	$a_n(2|2)$	$a_n(3|2)$	$a_n(4|2)$	$a_n(5|2)$	$a_n(6|2)$
	1	$b(1)$	$a_n(1|1)$	$a_n(2|1)$	$a_n(3|1)$	$a_n(4|1)$	$a_n(5|1)$	$a_n(6|1)$

Note—Rat., rating; Par., parameter. Old/new and source ratings range from 6 (*confident old* and *confident Source B*, respectively) to 1 (*confident new* and *confident Source A*, respectively).

Similarly, in mental state M_5, there is neither item memory nor distractor detection. Parameters $b(1), \ldots, b(6)$ for guessing the old/new response correspond conceptually to the b parameter of the dichotomous case. The source-rating parameters in state M_5, given an *old* response, have been named a_o, because they correspond to the source-guessing parameter a in the dichotomous case. Unlike in the dichotomous case, source ratings are now also available for *new* responses; hence, new parameters for source guessing are required conditional on the *new* responses in state M_5. They are named a_n.

Simplifying the State–Response Mapping

It is in the nature of discrete-state models that they require many parameters for complex response formats. If the interest is in measuring item and source memory, the parameters governing the state–response mapping can be considered nuisance parameters, and estimation and inference can focus on the parameters of the stimulus–state mapping. One advantage is that the memory parameters governing the stimulus–state mapping are then estimated with minimal assumptions about scale usage.[1] However, in a model comparison context, models with fewer parameters, such as the Hautus et al. (2008) models, would probably be preferred for reasons of parsimony.

The state–response mapping can be considerably simplified using substantive assumptions. The adequacy of the assumptions can be assessed by the simplified model's ability to fit the empirical data. In the present case, a standard move frequently made for dichotomous data is to set equal source-guessing parameters g and a (e.g., Bayen et al., 1996). This can analogously be done for rating data, and thus, we set equal corresponding a_o and g_o parameters, as well as corresponding a_n and g_n parameters.[2]

Furthermore, in states M_1 and M_2, there is source memory for A and B, respectively. We assumed that scale usage for source ratings would be symmetrical in this case, and thus, we constrained source-rating parameters s expressing equal levels of confidence to be equal for M_1 and M_2.[3]

In a similar vein, we assumed that scale usage for old/new ratings would be the same in these two states. Thus, we constrained the three old/new parameters r for state M_1 to be equal to those for state M_2.

Next, we used three psychological assumptions, pertaining to the so-called extreme response style, consistency between ratings, and a partial ordering of mental states according to level of confidence. The extreme response style refers to a finding from research on response styles. It is found in ratings requiring the participant to respond along an intensity dimension, such as level of confidence. In such situations, some persons have a tendency to use the extreme alternatives, whereas others tend to employ the middle categories with greater frequency (e.g., Hamilton, 1968). In the present context, this means that we can expect a correlation between rating extremity for old/new ratings and rating extremity for source ratings caused by interindividual differences in extreme response style.[4]

Consider the 6×6 table cross-tabulating old/new ratings and source ratings, with response frequencies aggregated across participants. Moving down from the row with

old response of highest confidence (i.e., 6) to rows with lower levels of confidence, interindividual differences in extreme response style should lead to dispersed data in the first row with many extreme source ratings, but source ratings should be successively more compressed toward the cautious middle source ratings (i.e., 4 or 3) for lower levels of confidence in the *old* response.

A second principle is response editing for consistency between ratings (Tourangeau, Rips, & Rasinski, 2000). Participants attempt to maintain consistency between successive ratings. For example, it would be inconsistent to respond *new* to a test item and then ascribe it to one of the two sources with high confidence. In other words, to the extent to which the item is judged new, source ratings should be compressed toward the middle categories (see also Slotnick & Dodson, 2005). This would add to the effects of interindividual differences in extreme response style for *old* responses and would counteract, and perhaps override, such effects for *new* responses. Hautus et al. (2008) appealed to a similar idea to justify the use of a separate set of guessing parameters for the *new* responses in their Models 2 and 3.

A third observation is that the mental states considered here differ in the amount of information that is available in each case. Thus, in states M_1 and M_2, there is both item and source information, whereas in M_3 and M_4, there is only item information but no source information. Finally, in M_5, there is not even item information. It seems plausible that this could lead to differences between these states in the overall level of confidence with which old/new ratings are made and, hence, in the willingness to use extreme response categories for these ratings. Similarly, there might be differences between the states with and without source memory in the overall level of confidence with which source ratings are made.

Each of these principles suggests that certain groups of parameters should differ mostly in the extent to which responses are compressed toward the middle categories. For example, the extreme response style and the consistency principle suggest that parameter values $g_o = a_o$ should be most dispersed across the six columns in the row with an old/new rating of 6 (see Table 1), but successively compressed toward the middle for old/new ratings of 5 and 4. That is, in these rows, the probability of cautious source ratings should increase at the expense of the probability of extreme source ratings. Response editing for consistency similarly suggests a compression toward the middle in source ratings given a *new* response for parameters $g_n = a_n$.

We modeled this idea by defining a compression function. The compression function used two parameters: a parameter for overall shape and a compression factor λ, regulating the amount of compression toward middle scale categories. Figure 2 shows how the function works. It takes as input a set of probabilities for the six rating categories and transforms these by compressing them toward the middle. In Figure 2, this is shown starting from a probability distribution with equal probabilities for the six rating categories (line labeled $\lambda = 0$). This is subjected to four transformations with an increasingly high compres-
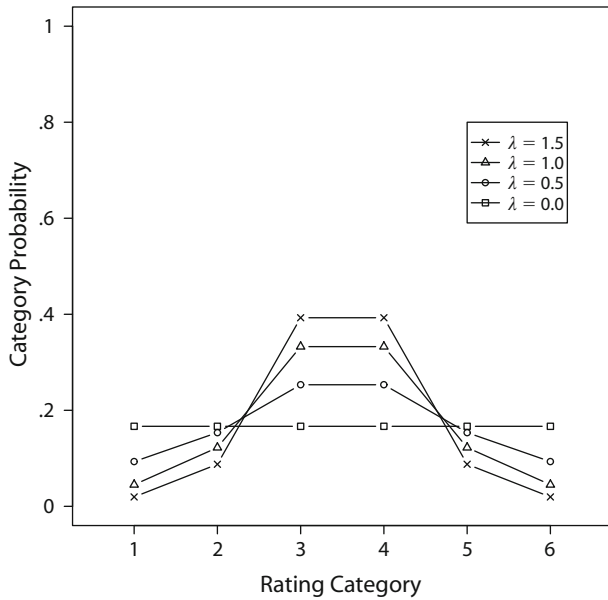
Figure 2. The compression function for four levels of compression λ.

sion factor of λ = 0.5, 1.0, and 1.5. It can be seen that the effect is to weight middle categories successively more heavily at the expense of the extreme categories.

Use of the compression function allows us to model most parameter groups more parsimoniously. For example, the parameters $g_o = a_o$ are modeled by using only one set of six parameter values (of which five are nonredundant) to describe source ratings under the highest level of old/new confidence—that is, for the row with an old/new rating of 6 in Table 1. The parameter values for rows with less extreme *old* ratings are generated from these through the use of two compression factors: one for the row with an old/new rating of 5 and one for the row with an old/new rating of 4. This replaces $15 = 3 \times 5$ parameters with $7 = 5 + 2$ parameters. Details are described in the Appendix. The resulting simplified model has 32 parameters for the Yonelinas (1999, Experiment 2) data and 36 parameters for the Slotnick et al. (2000, Experiments 2 and 3) data.

RESULTS

The discrete-state model was fitted to the three data sets analyzed by Hautus et al. (2008)—that is, to the data from Yonelinas (1999, Experiment 2) and the data from Slotnick et al. (2000, Experiments 2 and 3)—using the maximum likelihood method.[5] Table 2 reports the parameter estimates for the stimulus–state mapping. As can be seen, item memory parameters, D_A and D_B, and the parameters for source memory given item memory, d_A and d_B, were highest for Slotnick et al.'s Experiment 2.

Table 3 shows the estimated state–response mapping for the Yonelinas (1999) data in the format of Table 1. As can be seen, the s parameters for source rating (states M_1 and M_2), as well as the source-guessing parameters $g_o = a_o$ given *old* responses (states M_3 and M_5), exhibit the com-

pression pattern: Middle categories are increasingly preferred as confidence in the *old* response decreases. Source guessing given *new* responses (parameters $g_n = a_n$ for states M_4 and M_5) seems to be dominated by the consistency principle: After a *new* response, guesses are strongly biased toward the middle rating categories. Source guessing after *new* responses is, in fact, more or less independent of the level of confidence in the *new* response, indicating that there is room for further simplification of the state–response mapping.

To evaluate models, Hautus et al. (2008) presented graphical ROC analyses (for the Yonelinas, 1999, Experiment 2 data), as well as quantitative assessments of model performance. Figure 3 shows the observed ROCs and those predicted by the discrete-state model for the data in Yonelinas (1999, Experiment 2). The upper left panel presents the old/new ROCs pooled across source

Table 2
Estimates of Item and Source Memory Parameters

Data Set	D_A	D_B	D_N	d_A	d_B
Yonelinas (1999), Experiment 2	.43	.46	.32	.55	.55
Slotnick, Klein, Dodson, and Shimamura (2000), Experiment 2	.77	.76	.40	.63	.59
Slotnick et al. (2000), Experiment 3	.64	.64	.47	.50	.48

Table 3
State–Response Mapping Estimated for Data by
Yonelinas (1999, Experiment 2)

Mental State	Old/New Rat.	Old/New Par.	Source Rating Parameters					
			1	2	3	4	5	6
M_1	6	.90	.90	.10	.00			
	5	.09	.00	.96	.04			
	4	.00	.00	.00	1.00			
	3							
	2			need not be specified				
	1							
M_2	6	.90				.00	.10	.90
	5	.09				.04	.96	.00
	4	.02				1.00	.00	.00
	3							
	2			need not be specified				
	1							
M_3	6	.83	.16	.17	.10	.15	.22	.21
	5	.16	.07	.18	.18	.26	.23	.09
	4	.01	.01	.13	.27	.40	.17	.02
	3							
	2			need not be specified				
	1							
M_4	6							
	5			need not be specified				
	4							
	3	.09	.01	.02	.33	.51	.10	.02
	2	.42	.01	.02	.33	.52	.10	.02
	1	.49	.01	.03	.32	.50	.11	.03
M_5	6	.08	.16	.17	.10	.15	.22	.21
	5	.15	.07	.18	.18	.26	.23	.09
	4	.20	.01	.13	.27	.40	.17	.02
	3	.22	.01	.02	.33	.51	.10	.02
	2	.22	.01	.02	.33	.52	.10	.02
	1	.13	.01	.03	.32	.50	.11	.03

Note—Rat., rating; Par., parameter. Old/new and source ratings range from 6 (*confident old* and *confident Source B*, respectively) to 1 (*confident new* and *confident Source A*, respectively).
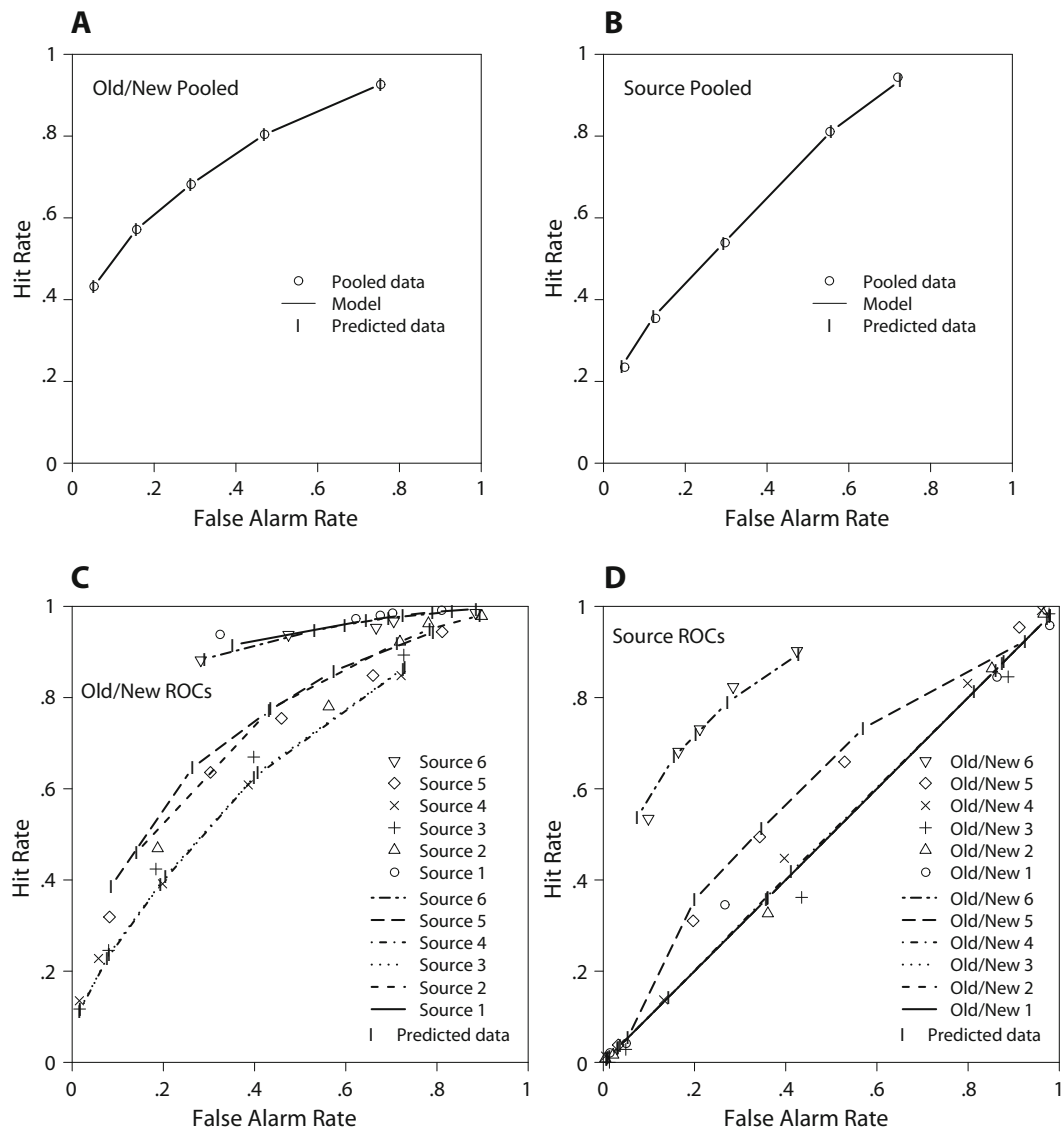
Figure 3. Model fit to the receiver-operating characteristic (ROC) data by Yonelinas (1999, Experiment 2). The left half of the figure shows old/new data and model predictions; the right half shows source data and predictions. The upper panels show pooled analyses; the lower panels refined analyses (see the text). Symbols indicate observed data, whereas bars and curves show the model predictions for the observed data points and the ROCs, respectively.

ratings; the upper right panel presents the source ROCs pooled across old/new ratings. As can be seen, the observed ROCs are almost perfectly fitted by the discrete-state model. The lower panels show so-called refined old/new and source ROCs (Hautus et al., 2008). There are six refined ROCs for each task, one for each level of confidence on the other task. For example, there are six refined old/new ROCs (lower left panel), one for each level of confidence in source ratings. As can be seen, the observed refined ROCs are fitted reasonably well, at a level comparable to the performance of Hautus et al.'s (2008) Model 3.

To evaluate models quantitatively, Hautus et al. (2008) used an array of goodness-of-fit and model selection indices (see Hautus et al., 2008, for details).

G^2 is a goodness-of-fit statistic; $G^2 = 2\Sigma O \log(O/E)$, where O and E are the observed and predicted frequencies, respectively, in the cells of the contingency tables cross-tabulating old/new responses and source responses. Smaller values indicate better fits.

RMSEA is another goodness-of-fit index, the so-called root-mean square error of approximation. Values of RMSEA smaller than 0.05 are considered an indication of close fit (see DeCarlo, 2003b, Appendix; Hautus et al., 2008).

AIC and BIC are the Akaike information criterion and the Bayesian information criterion, respectively. They are model selection indices that introduce penalties for a model's flexibility in terms of the model's

number of parameters. Models with smaller values are preferred, indicating that they strike a better compromise between fit and parsimony than do models with larger values (Pitt, Myung, & Zhang, 2002).[6]

Table 4 presents these statistics for the three models proposed by Hautus et al. (2008), for Onyper et al.'s (2010) model, and for the discrete-state model, separately for each of the three data sets analyzed by Hautus et al. As has already been mentioned, Hautus et al. accepted Model 3 because it outperformed all other models on all indices, with the exception of a better BIC value for Model 2 than for Model 3 for the Slotnick et al. (2000) Experiment 3 data (see Table 4). The discrete-state model outperforms Hautus et al.'s model with respect to all statistics, with the exception of a tie in BIC values with Model 2 for the Slotnick et al. Experiment 2 data and a slightly better BIC value for Model 3 for this data set. As compared with Onyper et al.'s model, the discrete-state model describes the Slotnick et al. Experiment 2 data better than does Onyper et al.'s model in terms of the model selection indices AIC and BIC, whereas Onyper et al.'s model describes the Yonelinas (1999) Experiment 2 data better than does the discrete-state model. Both models are tied for the Slotnick et al. Experiment 3 data, in that AIC favors Onyper et al.'s model, whereas BIC favors the discrete-state model.

DISCUSSION

Hautus et al.'s (2008) Model 3 is a hybrid model combining features of two-dimensional signal detection models with features of discrete-state models. Thus, there are two mental states and a rudimentary stimulus–state mapping: Targets can have been attended (State 1) or not (State 2). Given a target, the two states are entered into with mixture probability of p. In addition, there is response editing akin to a state–response mapping: Source responses following *old* responses are based on the signals provided by the continuous signal detection part of the model; source responses following *new* responses are based on guessing probabilities, as in a state of uncertainty in a discrete-state model. Similarly, Onyper et al.'s (2010) model involves two mental states and a rudimentary stimulus–state mapping: Diagnostic item and source information can be retrieved for a target (State 1: recollection), or there can be only item information (State 2: familiarity). In addition, there is a relatively complex response mapping in that a separate set of response bounds for source judgments is used for each level of old/new rating. In the present article, we went all the way to a discrete-state model. We then placed the model in the set of models considered by Hautus et al. and tested how it performed according to the rules adopted by these authors. In terms of these rules, the present model wins the competition over the models considered by Hautus et al. As compared with Onyper et al.'s model, the model comparison resulted in a tie.

We believe that the discrete-state model is thereby established as a viable alternative to the model endorsed by Hautus et al. (2008) and to Onyper et al.'s (2010) model. One limitation of these analyses is that the model selection criteria AIC and BIC make only rough and approximate corrections for a model's flexibility (Pitt et al., 2002). More appropriate corrections would take the functional form of the model into account, as is done in model selection criteria based on minimum description length (Pitt et al., 2002). Although some progress has been made in computing such criteria for multinomial models (Wu, Myung, & Batchelder, 2010), their computation is a formidable task, beyond the scope of the present article, for the set of models considered here. In consequence, it is possible that some of the competing models, the signal detection models or the

Table 4
Fit Indices and Model Selection Indices

	Yonelinas (1999, Experiment 2)					Slotnick, Klein, Dodson, and Shimamura (2000, Experiment 2)					Slotnick et al. (2000, Experiment 3)				
p/df	G^2	RMSEA	AIC	BIC		p/df	G^2	RMSEA	AIC	BIC	p/df	G^2	RMSEA	AIC	BIC
						DeCarlo's (2003b) Model									
19/86	2,418	0.119	2,456	2,507		21/123	1,582	0.117	1,624	1,687	21/123	3,765	0.124	3,807	3,870
						Hautus, Macmillan, and Rotello's (2008) Model 1									
19/86	951	0.072	989	1,040		21/123	620	0.068	662	725	21/123	1,186	0.067	1,228	1,291
						Hautus et al.'s (2008) Model 2									
24/81	213	0.029	261	326		27/117	199	0.029	253	334	27/117	212	0.021	266	347
						Hautus et al.'s (2008) Model 3									
29/76	178	0.026	236	314		32/112	173	0.025	237	333	32/112	193	0.019	257	353
						Onyper, Zhang, and Howard's (2010) Model									
40/65	94	0.015	174	281		53/91	140	0.025	246	405	53/91	107	0.010	213	371
						Discrete-State Model									
32/73	159	0.025	223	308		36/108	155	0.022	227	334	36/108	162	0.016	234	342

Note—p/df, number of parameters/degrees of freedom; G^2, goodness-of-fit statistic; RMSEA, a measure of approximate fit (RMSEA < .05 indicates close fit); AIC, Akaike information criterion computed from G^2 values by adding $2p$; BIC, Bayesian information criterion computed from G^2 by adding $p\log(n)$, where n is the number of cells in the data matrix (see Hautus et al., 2008). Values of AIC and BIC for Hautus et al.'s models were recomputed for the Slotnick et al. (2000) data sets because some of the values in Hautus et al. (Table 3) are not consistent with the reported G^2 values; see also note 5.

discrete-state model, would be penalized more strongly for lack of parsimony than is apparent in AIC or BIC when functional form is taken into account. Nevertheless, the good overall fit of the discrete-state model suggests that the data sets analyzed here do not support a firm decision between the discrete-state model, the model finally accepted by Hautus et al., and Onyper et al.'s model.

In modeling complex sets of data, it is almost inevitable that some of the design choices in building a model are ad hoc, data driven, or made out of mathematical convenience, and Hautus et al. (2008) openly acknowledged this in developing their models, where appropriate.[7] The major ad hoc choice in the discrete-state model is the particular function chosen to model the compression in the state–response mapping. The particular exponential function that we used was our first and only choice, but many different functional forms could have been tried out. It seems likely that playing around with different compression functions might sometimes improve and sometimes degrade the model's performance in terms of the criteria used to evaluate models.

One advantage of the discrete-state model is that it uses the same model for the stimulus–state mapping as does Bayen et al.'s (1996) model for dichotomous data, irrespective of response formats. The measurement of memory performance in terms of item and source memory parameters is therefore straightforward and conceptually clear, irrespective of response format. This reflects the fact that different response formats do not alter the stimulus–state mapping. Only the state–response mapping must be adapted.

The downside of this is that major portions of the data were modeled by the state–response mapping. In doing so, we used three substantive assumptions: (1) interindividual differences (and possibly intraindividual differences; see note 4) in extreme response style; (2) response editing for consistency; and (3) the idea that different mental states may differ in overall level of confidence. Assumptions 1 and 2 are based on well-documented phenomena (e.g., Hamilton, 1968; Tourangeau et al., 2000), and Assumption 3 seems reasonable.

Nevertheless, it could be argued that Hautus et al.'s (2008) Model 3 is the more principled model, inasmuch as a major pattern in the present data sets flows naturally from its assumption. The major pattern is that source ratings are compressed toward the middle as *old* responses decrease in confidence and for *new* responses. This pattern is very strong in the Slotnick et al. (2000) data in particular, and it flows from the combination of several features of Hautus et al.'s model: (1) The bivariate correlations of the target distributions, leading to reduced source discrimination for targets with low values on the old/new dimension; (2) the use of likelihood ratio bounds for source decisions; these bounds spread out as a target's value on the old/new dimension decreases from values typical for the target toward values typical for the lures (see, e.g., Figure 8 in Hautus et al., 2008); and (3) the use of a distribution for unattended items, leading to lower levels of source discrimination as old/new confidence decreases (Hautus et al., 2008, p. 900).[8] In contrast, in the discrete-state model, the compression pattern

is modeled entirely by the state–response mapping in what could be argued to be ad hoc simplifications of it, and there is some truth to that argument.

But the argument can also be turned around. Response editing for consistency and variations in extreme response style by themselves account for the observed compression pattern. These phenomena modulate scale usage, and hence, it is appropriate to model their effects in the state–response mapping of the model. Conversely, where these phenomena are not factored into the response-mapping part of a model, they will artifactually distort other parts of the model. Thus, it could be argued that the pattern of bivariate correlations, the use of likelihood ratio bounds, and the addition of a noise distribution are distortions forced upon the signal detection framework to cope with a compression pattern that, in reality, largely reflects scale-usage phenomena. In line with this argument, the potentially distorting role of response processes is currently receiving increased attention in the signal detection literature (Benjamin, Diaz, & Wee, 2009; Mueller & Weidemann, 2008).

There are two conclusions to be drawn from the present modeling exercise. First, the analyzed data are not strong enough to allow one to discriminate between the kind of hybrid signal detection model proposed by Hautus et al. (2008) and Onyper et al. (2010), on the one hand, and discrete-state models, on the other hand. New and structurally enriched data are required to make this discrimination. Several avenues of research seem promising. A simple approach is to go back to dichotomous *old* versus *new* responses and dichotomous Source A versus Source B responses and to manipulate decision bias in item and source recognition experimentally, rather than via confidence ratings. Base rate and payoff manipulations could be used for the purpose (Van Zandt, 2000). Signal detection models and discrete-state models make unambiguously different predictions for such experiments (Bröder & Schütz, 2009).

A second approach, independent of response format, is to introduce experimental manipulations targeting certain model parameters of one of the models. For example, a manipulation might be implemented that can be expected, on the basis of substantive psychological theory and independent findings, to influence selectively the proportion p of targets not attended, as per Hautus et al.'s (2008) Model 3. The effect of this manipulation should then map onto the appropriate parameter p of that model selectively, if the model is true, and it would in all likelihood lead to less parsimonious changes in the model parameters of the discrete-state model. Such a result pattern would support the Hautus et al. model and subtract from the plausibility of the discrete-state model.

A third approach in this spirit is to test for source memory for unrecognized items—that is, for evidence for source discrimination among items receiving *new* ratings. Neither Hautus et al.'s (2008) Model 3 nor the present discrete-state model could account for such a finding without modification. Both in Hautus et al.'s model and in the discrete-state model, source ratings for items rated *new* are guesses. But as was pointed out by Starns, Hicks, Brown, and Martin (2008), source discrimination

for unrecognized items would flow naturally from a bivariate signal detection model without source guessing for *new* responses under certain conditions, and Onyper et al.'s (2010) model could probably model it. There is, however, little evidence for source memory for unrecognized targets in either the Slotnick et al. (2000) data or the Yonelinas (1999) data analyzed here.

Starns et al. (2008) reported three experiments in which they found source discrimination for unrecognized targets in a condition with a conservative old/new criterion. As was acknowledged by Starns et al., the results from the first two experiments are open to alternative explanations, due to the fact that old/new decisions and source ascriptions were collected in separate phases of the experiment. In Experiment 3, item and source decisions were made successively for each item presented at test. However, in this experiment, an unusual feature was that participants knew that an unrecognized target was an old item when they were asked to ascribe it to a source despite an initial *new* response. Knowledge of this may have promoted a second attempt at retrieving information for the item, an attempt that may sometimes have elicited sufficient information to allow the participants to ascribe the item to the correct source (and to correct the initial *new* judgment if the participants had been allowed to do so). In sum, although it is suggestive and worth further research, we feel that the evidence on source recognition for unrecognized items is currently not strong enough to stop thinking about discrete-state models.

A second conclusion is a caveat. As was pointed out above, interindividual differences in scale usage may be responsible for some of the systematic patterns in the data analyzed here, and they have the potential to distort one's model when not taken into account. More generally, a model that fits each individual's data need not fit the data aggregated across individuals if participants differ in scale usage or memory performance. Conversely, a model fitting the aggregate data may well be outside the class of models that generate each individual's data. Interindividual differences are ubiquitous, and they need to be taken into account if any progress is to be made in uncovering the processes underlying participants' responses.

Individual data are usually not rich enough to fit the kind of model considered here for each individual separately, but there is a compromise between the extremes of individual and aggregate analyses, the hierarchical modeling approach (Raudenbush & Bryk, 2002). Considerable progress has been made in the last decade in developing hierarchical models for discrete-state models (Klauer, 2006, 2010; Smith & Batchelder, 2010), as well as for signal detection models (Rouder & Lu, 2005), and we believe that the field would profit from using such models to make appropriate corrections for the distorting effects of interindividual differences in the analyses.

In conclusion, what are the implications of the present analyses and model for memory theory? Conceptually, the three distributions for targets in Onyper et al.'s (2010) model—one for Source A items with diagnostic item and source information, one for Source B items with diagnostic item and source information, and one for studied items with diagnostic item information, but without source information—correspond, in order, to the three states M_1 (item and source memory for Source A items), M_2 (item and source memory for Source B items), and M_3 (item memory but no source memory) of the discrete-state model. Both models are thus in line with the assumption that recognition memory depends on two distinguishable memory signals, often termed *familiarity* and *recollection*, which they conceptualize in roughly similar ways as processes subserving item memory and item-cum-source-memory, respectively. They thereby provide measurement tools for quantifying the separate contributions of these processes in observed recognition data (Batchelder & Riefer, 1999), and they impose constraints on the development of more fine-grained models of the underlying encoding and retrieval processes (e.g., Shiffrin & Steyvers, 1997). Both models also agree in that states with diagnostic information are not simply mapped onto highest confidence responses but always give rise to a range of responses.

There are, however, two major theoretical differences between the discrete-state model and signal detection models. One is the obvious difference that memory signals are graded along continuous dimensions in the signal detection models, whereas they are all-or-none in the discrete-state model. The second, less obvious, difference is that the discrete-state model incorporates a state of distractor detection, a process of active rejection for new items, that has no conceptual counterpart in the signal detection models. As was already mentioned, distractor detection is assumed to be based on metacognitive inference. For example, seeing a new item, participants may infer that it is new on the basis of, for example, some unusual feature of it, idiosyncratic associations it evokes, or its overall incongruency with the set of targets (Strack & Bless, 1994), among other things. Supplementing the strategies for discriminating between models considered above, distractor detection, in particular, may well turn out to be a source of several distinctive predictions of discrete-state models.

REFERENCES

ASHBY, F. G. (1988). Estimating the parameters of multidimensional signal detection theory from simultaneous ratings on separate stimulus components. *Perception & Psychophysics*, **44**, 195-204.

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

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

BATCHELDER, W. H., & 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 processing tree modeling. *Psychonomic Bulletin & Review*, **6**, 57-86.

BATCHELDER, W. H., RIEFER, D. M., & HU, X. (1994). Measuring memory factors in source monitoring: Reply to Kinchla. *Psychological Review*, **101**, 172-176.

BAYEN, U. J., MURNANE, K., & ERDFELDER, E. (1996). Source discrimination, item detection, and multinomial models of source monitoring. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **22**, 197-215.

BAYEN, U. J., NAKAMURA, G. V., DUPUIS, S. E., & YANG, C.-L. (2000). The use of schematic knowledge about sources in source monitoring. *Memory & Cognition*, **28**, 480-500.

BENJAMIN, A. S., DIAZ, M. L., & WEE, S. (2009). Signal detection with criterion noise: Applications to recognition memory. *Psychological Review*, **116**, 84-115.

BROADBENT, D. E. (1966). Two-state threshold model and rating-scale experiments. *Journal of the Acoustical Society of America*, **40**, 244-245.

BRÖDER, A., & SCHÜTZ, J. (2009). Recognition ROCs are curvilinear—or are they? On premature arguments against the two-high-threshold model of recognition. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **35**, 587-606.

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. (2003a). An application of signal detection theory with finite mixture distributions to source discrimination. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **29**, 767-778.

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

DODSON, C. S., HOLLAND, P. W., & SHIMAMURA, A. P. (1998). On the recollection of exact and partial source information. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **24**, 1121-1136.

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

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

FALMAGNE, J.-C. (1985). *Elements of psychophysical theory*. New York: Oxford University Press.

GLANZER, M., HILFORD, A., & KIM, K. (2004). Six regularities of source recognition. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **30**, 1176-1195.

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

HAMILTON, D. L. (1968). Personality attributes associated with extreme response style. *Psychological Bulletin*, **69**, 192-203.

HAUTUS, M. J., MACMILLAN, N. A., & ROTELLO, C. M. (2008). Toward a complete decision model of item and source recognition. *Psychonomic Bulletin & Review*, **15**, 889-905.

HILFORD, A., GLANZER, M., KIM, K., & DECARLO, L. T. (2002). Theories of source memory: ROC analysis. *Journal of Experimental Psychology: General*, **131**, 494-510.

JOHNSON, M. K., HASHTROUDI, S., & LINDSAY, D. S. (1993). Source monitoring. *Psychological Bulletin*, **114**, 3-28.

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

JOHNSON, M. K., & RAYE, C. L. (1981). Reality monitoring. *Psychological Review*, **88**, 67-85.

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

KLAUER, K. C. (2006). Hierarchical multinomial processing tree models: A latent-class approach. *Psychometrika*, **71**, 1-31.

KLAUER, K. C. (2010). Hierarchical multinomial processing tree models: A latent-trait approach. *Psychometrika*, **75**, 70-98.

KLAUER, K. C., EHRENBERG, K., & WEGENER, I. (2003). Crossed categorization and stereotyping: Structural analyses, effect patterns, and dissociative effects of context relevance. *Journal of Experimental Social Psychology*, **39**, 332-354.

KLAUER, K. C., & WEGENER, I. (1998). Unraveling social categorization in the "Who said what?" paradigm. *Journal of Personality & Social Psychology*, **75**, 1155-1178.

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

LARKIN, W. D. (1965). Rating scales in detection experiments. *Journal of the Acoustical Society of America*, **37**, 748-749.

LEE, W. (1963). Choosing among confusably distributed stimuli with specified likelihood ratios. *Perceptual & Motor Skills*, **16**, 445-467.

MACMILLAN, N. A., & CREELMAN, C. D. (2005). *Detection theory: A user's guide* (2nd ed.). Mahwah, NJ: Erlbaum.

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

MALMBERG, K. J. (2008). Recognition memory: A review of the critical findings and an integrated theory for relating them. *Cognitive Psychology*, **57**, 335-384.

MEISER, T. (2005). A hierarchy of multinomial models for multidimensional source monitoring. *Methodology*, **1**, 2-17.

MEISER, T., & BRÖDER, A. (2002). Memory for multidimensional source information. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **28**, 116-137.

MEISER, T., SATTLER, C., & VON HECKER, U. (2007). Metacognitive inferences in source memory judgements: The role of perceived differences in item recognition. *Quarterly Journal of Experimental Psychology*, **60**, 1015-1040.

MUELLER, S. T., & WEIDEMANN, C. T. (2008). Decision noise: An explanation for observed violations of signal detection theory. *Psychonomic Bulletin & Review*, **15**, 465-494.

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

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

NACHMIAS, J., & STEINMAN, R. M. (1963). Study of absolute visual detection by the rating-scale method. *Journal of the Optical Society of America*, **53**, 1206-1213.

ONYPER, S. V., ZHANG, Y., & HOWARD, M. W. (2010). Some-or-none recollection: Evidence from item and source memory. *Journal of Experimental Psychology: General*, **139**, 341-364.

PITT, M. A., MYUNG, I. J., & ZHANG, S. (2002). Toward a method of selecting among computational models of cognition. *Psychological Review*, **109**, 472-491.

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.

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

RAUDENBUSH, S. W., & BRYK, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*. Thousand Oaks, CA: Sage.

R DEVELOPMENT CORE TEAM (2009). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.

RIEFER, D. M., HU, X., & BATCHELDER, W. H. (1994). Response strategies in source monitoring. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **20**, 680-693.

ROUDER, J. N., & LU, J. (2005). An introduction to Bayesian hierarchical models with an application in the theory of signal detection. *Psychonomic Bulletin & Review*, **12**, 573-604.

SHIFFRIN, R. M., & STEYVERS, M. (1997). A model for recognition memory: REM—Retrieving effectively from memory. *Psychonomic Bulletin & Review*, **4**, 145-166.

SLOTNICK, S. D., & DODSON, C. S. (2005). Support for a continuous (single-process) model of recognition memory and source memory. *Memory & Cognition*, **33**, 151-170.

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, & Cognition*, **26**, 1499-1517.

SMITH, J. B., & BATCHELDER, W. H. (2010). Beta-MPT: Multinomial processing tree models for addressing individual differences. *Journal of Mathematical Psychology*, **54**, 167-183.

SPENCER, W. D., & RAZ, N. (1995). Effects of aging on memory for content and context: A meta-analysis. *Psychology & Aging*, **10**, 527-539.

STAHL, C., & KLAUER, K. C. (2007). HMMTree: A computer program for latent-class hierarchical multinomial processing tree models. *Behavior Research Methods*, **39**, 267-273.

STARNS, J. J., HICKS, J. L., BROWN, N. L., & MARTIN, B. A. (2008). Source memory for unrecognized items: Predictions from multivariate signal detection theory. *Memory & Cognition*, **36**, 1-8.

STRACK, F., & BLESS, H. (1994). Memory for nonoccurrences: Metacognitive and presuppositional strategies. *Journal of Memory & Language*, **33**, 203-217.

TANNER, W. P., JR. (1956). Theory of recognition. *Journal of the Acoustical Society of America*, **28**, 882-888.

TOURANGEAU, R., RIPS, L. J., & RASINSKI, K. (2000). *The psychology of survey response*. Cambridge: Cambridge University Press.

VAN ZANDT, T. (2000). ROC curves and confidence judgments in recognition memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **26**, 582-600.

WATSON, C. S., & BOURBON, W. T. (1965). Rating scales and two-state threshold models. *Journal of the Acoustical Society of America*, **38**, 667-668.

WICKELGREN, W. A. (1968). Testing two-state theories with operating characteristics and a posteriori probabilities. *Psychological Bulletin*, **69**, 126-131.

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

WIXTED, J. T. (2007). Dual-process theory and signal-detection theory of recognition memory. *Psychological Review*, **114**, 152-176.

WU, H., MYUNG, J. I., & BATCHELDER, W. H. (2010). Minimum description length model selection of multinomial processing tree models. *Psychonomic Bulletin & Review*, **17**, 275-286.

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

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

YONELINAS, A. P., & PARKS, C. M. (2007). Receiver operating characteristics (ROCs) in recognition memory: A review. *Psychological Bulletin*, **133**, 800-832.

NOTES

1. A few simplifications are, however, necessary to make the model identified.

2. More precisely, $g_o(y|x)$ was constrained to equal $a_o(y|x)$ for $y = 1, \ldots, 6$ and $x = 4, 5$, and 6, and $a_n(y|x)$ was constrained to equal $g_n(y|x)$ for all $y = 1, \ldots, 6$ and $x = 1, 2$, and 3.

3. More precisely, $s(y|x)$ was constrained to equal $s(7 - y|x)$ for $y = 1, 2$, and 3 and $x = 4, 5$, and 6.

4. Although the correlation is one between participants, it can be speculated that extreme response style, like other personality traits, also has a state component that leads to fluctuations in extreme response style from moment to moment within each participant. If so, the same correlation would also exist, perhaps at a reduced level, within each participant across trials.

5. The analyses were implemented in R (R Development Core Team, 2009). The R scripts can be obtained from the first author.

6. Hautus et al. (2008) computed BIC by adding $p\log(n)$ to G^2, where p is the number of nonredundant parameters in the model and n is the number of cells in the data matrix (see the note to their Table 3), and we followed them in defining BIC in that way. It is, however, more common to define BIC using some measure of sample size for n, such as the number of participants or the total number of responses, depending on what is considered the unit of analysis (e.g., Stahl & Klauer, 2007).

7. For example, likelihood ratio bounds are used only for the source decision, but not for the item decision, due to mathematical convenience. When a noise distribution for unattended targets is added, as in Hautus et al.'s (2008) Model 3, the likelihood ratio bounds are, in fact, no longer likelihood ratio bounds relative to the modified likelihood, making these bounds appear ad hoc. A data-driven choice is to consider old/new ratings of level 4 as *new* responses in the data in Slotnick et al. (2000), as acknowledged by Hautus et al. (p. 905). For Onyper et al.'s (2010) model, the choice of different response bounds for the source judgment as a function of the old/new response appears to be an ad hoc choice.

8. In Onyper et al.'s (2010) model, the compression pattern is handled by (1) the use of a separate distribution for targets with familiarity information but without recollection, which has an effect similar to that of adding a distribution for unattended items in Hautus et al.'s (2008) Model 3, and (2) the use of a separate set of response bounds for the source judgment for each old/new judgment. The estimated bounds spread out as the target's value on the old/new dimension decreases (see Figure 6 in Onyper et al., 2010).

APPENDIX
The Simplified State–Response Mapping

The Compression Function

Extremity of a rating category was quantified as the distance of category i from the midpoint m of the scale (e.g., $m = 3.5$ for a scale ranging from 1 to 6)—that is, as $|i - m|$. A shape parameter $u > 0$ regulates the subjective spacing of the rating categories; extremity, in fact, enters the compression function as $|i - m|^u$. For $u = 1$, the extremity scores increase linearly as i moves out from the center of the scale. For larger u, the increase is positively accelerated, and thus, the distances between adjacent categories increase as i moves out from the center.

Starting out from a probability distribution $\pi_1, \ldots, \pi_6$ on the rating scale (assuming a 6-point scale), a compressed version of it is generated by weighting each π_i with weight factor $w_i(\lambda)$ given by

$$w_i(\lambda) = \exp(-|i - m|^u \lambda), \qquad i = 1, \ldots, 6. \tag{A1}$$

The compressed probability distribution $\pi_1^*(\lambda), \ldots, \pi_6^*(\lambda)$ is then obtained by normalizing the weighted probabilities so that they sum to one. This is achieved by dividing each weighted probability by the sum of the weighted probabilities:

$$\pi_i^*(\lambda) = \frac{w_i(\lambda)\pi_i}{\sum_{j=1}^{6} w_j(\lambda)\pi_j}. \tag{A2}$$

The compression factor λ regulates the amount of compression; its effect is shown in Figure 2. Changes in parameters u and λ actually have similar effects, and we used parameter u sparingly, modeling most compressions via the compression factor λ.

Using the Function in the State–Response Mapping

Consider first the source-rating parameters. As was already mentioned, we set equal corresponding g_o and a_o parameters, as well as corresponding g_n and a_n parameters. We refer to these parameter groups as g_o and g_n, respectively, in the following. The same u parameter was used for both parameter groups. On the basis of the ideas of extreme response style and response editing for consistency, the row of g_o parameters with *old* rating 6 (highest *old* confidence; see Table 1) was not compressed ($\lambda = 0$), and the other two rows used the same parameters for the six source-rating categories, but transformed with different compression factors, one for each row. These same compression factors were also used for the s parameters that describe source ratings in states M_1 and M_2 (see Table 1). On the basis of the idea that states with source memory (M_1 and M_2) may differ in overall level of confidence for source ratings from states without it (M_3, M_4, and M_5), a different parameter u was, however, used for the s parameters as a group.[A1]

Extreme response style and the consistency principle also operate on the g_n parameters for source guessing, given a *new* response in states M_3 and M_4. Different compression factors were therefore used for the different rows of g_n parameters, one of which could be set to a zero value of no compression to fix the scale.

Turning to the old/new parameters, we set $u = 1$ for all of them. This can be done without loss of generality, because changes in u can be compensated for by changes in other parameters in this part of the model. As was already mentioned in the body of the text, the parameters r for *old* ratings in states M_1 and M_2 were set equal. These and the b parameters were estimated without compression factor ($\lambda = 0$), again without loss of generality. The same parameters r were also used for *old* ratings in state M_3 and for *new* ratings in state M_4 (mirrored at the scale's midpoint), but in line with the idea that M_3 and M_4 might differ in overall level of confidence for old/new rating from M_1 and M_2, different compression factors were used for the group of r parameters in state M_3 and for the group of r parameters in state M_4.

To summarize, the state–response mapping for the Yonelinas (1999) data comprises the following parameters:

1. Five (nonredundant) category parameters and two compression factors generating the g_o parameters. The category parameter estimates are the six source-rating parameters shown in Table 3 in the row with old/new rating 6 for state M_3 (and M_5), one of which is redundant. The category parameters in rows with old/new ratings 5 and 4 were generated from these via the compression function with compression factors λ estimated as 0.31 and 0.72, respectively.

2. Five category parameters and two compression factors for the g_n parameters. The category parameter estimates are the six source-rating parameters shown in Table 3 in the row with old/new rating 3 for state M_4 (and M_5), one of which is redundant. The category parameters in rows with old/new ratings 2 and 1 were generated from these, with compression factors λ estimated as 0.02 and -0.06, respectively.

3. One u parameter common to the g_o and g_n parameters. Its value was estimated to be 1.77.

4. Two (nonredundant) category parameters and one parameter u for the s parameters for source ratings in states M_1 and M_2. The category parameter estimates are the three parameter values shown in Table 3 in the row with old/new rating 6 for state M_1, one of which is redundant. The u parameter was estimated to be 8.15.

5. Five (nonredundant) b parameters for old/new ratings in state M_5. The parameter estimates are shown in Table 3 in the column headed "Old/New Parameters" for state M_5; one of them is redundant.

6. Two (nonredundant) category parameters and two compression factors for the r parameters for old/new ratings in states M_1, M_2, M_3, and M_4. The category parameter estimates are shown in Table 3 in the column headed "Old/New Parameters" for state M_1 (and M_2), one of which is redundant. The category parameters for states M_3 and M_4 (mirror imaged) were generated from these with compression factors λ estimated as 0.62 and 2.11, respectively.

Taken together, there are 27 parameters defining the simplified state–response mapping. The estimates for the five parameters of the stimulus–state mapping are shown in Table 2.

The same state–response mapping was used for the Slotnick and Dodson (2005) data with straightforward changes to accommodate for the 7-point rating scales used for these data. Following Hautus et al. (2008), the old/new categories 7, 6, and 5 were considered *old* categories, the remaining categories *new* categories for the source-guessing parameters g_o and g_n. Given item and/or source memory, we assumed that participants would use one of the three appropriate response categories on the respective rating scales, leaving out the middle category 4 (e.g., participants remembering that an item is old would choose one of the old/new categories 5, 6, or 7, as governed by the r parameters). The simplified state–response mapping requires 31 parameters for this response format.

APPENDIX NOTE

A1. We could have estimated new compression factors λ for these parameters, but this would have introduced two new parameters (one parameter λ for each row of *old* response confidence, one of which can be fixed to zero). As was already mentioned, parameter u has effects qualitatively similar to those of parameter λ, and it allowed us to model a different overall level of compression for the s parameters through the use of only one additional parameter.