

## OBSERVATIONS

# On the Form of ROCs Constructed From Confidence Ratings

Kenneth J. Malmberg  
Indiana University Bloomington

A classical question for memory researchers is whether memories vary in an all-or-nothing, discrete manner (e.g., stored vs. not stored, recalled vs. not recalled), or whether they vary along a continuous dimension (e.g., strength, similarity, or familiarity). For yes–no classification tasks, continuous- and discrete-state models predict nonlinear and linear receiver operating characteristics (ROCs), respectively (D. M. Green & J. A. Swets, 1966; N. A. Macmillan & C. D. Creelman, 1991). Recently, several authors have assumed that these predictions are generalizable to confidence ratings tasks (J. Qin, C. L. Raye, M. K. Johnson, & K. J. Mitchell, 2001; S. D. Slotnick, S. A. Klein, C. S. Dodson, & A. P. Shimamura, 2000, and A. P. Yonelinas, 1999). This assumption is shown to be unwarranted by showing that discrete-state ratings models predict both linear and nonlinear ROCs. The critical factor determining the form of the discrete-state ROC is the response strategy adopted by the classifier.

Qin, Raye, Johnson, and Mitchell (2001), Slotnick, Klein, Dodson, and Shimamura (2000), and Yonelinas (1999) attempted to distinguish between threshold (or discrete-state) and continuous-state models of memory by analyzing the form of receiver operating characteristics (ROCs) constructed from confidence ratings (Green & Swets, 1966). The authors assume that *double high-threshold* detection models predict linear ratings ROCs, and that continuous-state detection models predict nonlinear ratings ROCs (Qin et al., 2001, p. 1110; Slotnick et al., 2000, p. 1502; Yonelinas, 1999, p. 1419). These ratings models can be distinguished on the basis of the linearity of the ratings ROC only to the extent that these assumptions hold. For example, if a double high-threshold model can predict nonlinear ratings ROCs then the form of the empirical ROC may not be informative.

The discussion about the nature of detection in the current memory literature (e.g., Batchelder & Riefer, 1990; Batchelder, Riefer, & Hu, 1994; Glanzer, Kim, Hilford, & Adams, 1999; Kinchla, 1994; Slotnick et al., 2000; Yonelinas, 1999) is similar to one that took place in the perception literature in the 1960's. At that time, Broadbent (1966), Krantz (1969), Larkin (1965), and Nachmias and Steinman (1963) showed that *low-threshold* models could predict nonlinear ratings ROCs that are indistinguishable from those predicted by continuous-state models. In their summary of this research, Lockhart and Murdock (1970) warned that, "examination of the operating characteristic is, in general, inconclusive, especially if generated by confidence ratings" (p. 105).

---

This research was supported in part by National Institute of Mental Health National Research Service Award Postdoctoral Fellowship MH12643-01. I thank the members of my dissertation committee at the University of Maryland: Jim Burnes, Arie Kruglanski, Kevin Murnane, Tom Nelson, Don Perlis, and Dana Plude. I also thank Amy Criss, Bill Estes, and Rich Shiffrin for their comments on a previous version of this article.

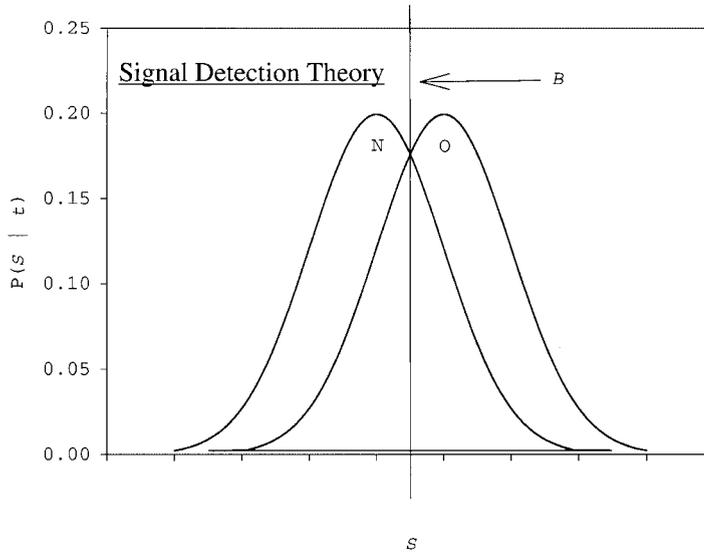
Correspondence concerning this article should be addressed to Kenneth J. Malmberg, Department of Psychology, Indiana University, Bloomington, Indiana 47405. E-mail: malmberg@indiana.edu

More recently, Erdfelder and Buchner (1998) showed how a threshold model of the inclusion–exclusion recognition task predicts nonlinear ratings ROCs (also see Buchner, Erdfelder, & Vaterrodt-Plunnecke, 1995, and Yonelinas & Jacoby, 1996). In this paper, I show, using similar but more general arguments, that the form of the ratings ROC predicted by a double-high threshold model, like those assumed by Qin et al. (2001), Slotnick et al. (2000), and Yonelinas (1997, 1999), is dependent on the assumptions concerning the mapping of the internal states to ratings. Before doing that, brief introductions are given to continuous- and discrete-state models, and then the implications of different methods for constructing ROCs are described.

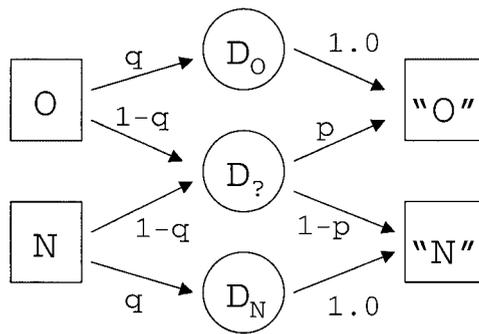
### Continuous-State Models—Signal Detection Theory

The task of a person performing a binary classification task is to decide to which one of two mutually exclusive and exhaustive classes a stimulus,  $t$ , belongs. Perhaps the most widely applied continuous-state theory of binary classification is signal detection theory (SDT; Green & Swets, 1966). SDT assumes that stimuli can vary on many dimensions, but the assignment of items to classes is based on the degree to which items are characterized by some property,  $S$ , as measured on a continuous, single-dimensional scale (Green & Swets, 1966). The top panel of Figure 1 illustrates the classical equal-variance SDT model, which assumes that  $S$  is a normally distributed random variable.

Classification will be successful to the extent that members of the different stimulus classes can be discriminated from each other on the basis of the degree to which items share similar values of  $S$ . One class of stimuli will be referred to as  $O$ , and the other class will be referred to as  $N$ . Within the framework of classical SDT (Green & Swets, 1966), the sensitivity with which the classifier can discriminate between  $O$  and  $N$  items,  $d'$ , is conceptualized as the normalized distance between the mean of the  $N$  and  $O$  probability-density distributions of  $S$  (assuming the  $N$  and  $O$  distributions are Gaussian, and  $\sigma_N = \sigma_O$ ),



Double High-Threshold Theory



		state		
		D <sub>O</sub>	D <sub>?</sub>	D <sub>N</sub>
stimulus	O	q	1-q	0
	N	0	1-q	q
		response		
		"O"	"N"	
state	D <sub>O</sub>	1.0	0	
	D <sub>?</sub>	p	1-p	
	D <sub>N</sub>	0	1.0	

Figure 1. Equal-variance, signal detection, and double high-threshold yes-no models.

$$d' = \frac{\mu_O - \mu_N}{\sigma_N} \tag{1}$$

Thus, discrimination is relatively sensitive when the difference between the average scalar value of the members of *O* and *N* is relatively large, and there is little variability in *S*.

For the yes-no classification task, it is assumed the decision rule partitions the classification scale at some point, called the *criterion*. One way to do so is based on the likelihood ratio,  $l(S)$ ,

$$l(S) = \frac{P(S|O)}{P(S|N)} \tag{2}$$

SDT is a continuous-state theory because  $l(S)$  is a continuous variable.

Decision rules in a likelihood-ratio model involve comparing  $l(S)$  with the criterion,  $\beta$ . If  $l(S) \geq \beta$ , then  $t$  is assigned to *O*, and

otherwise it is assigned to *N* (Green & Swets, 1966). Increasing the criterion causes fewer items to be classified as *O*, and decreasing the criterion causes more items to be classified as *O*. Because *S* is distributed normally, however, it is not the case that only *O* items have the potential to surpass the criterion. Likewise, it is never the case that only *N* items do not have the potential to surpass the criterion.

*Discrete-State Models*

Discrete-state theories assume that the presentation of a stimulus results in the classifier being in one state of a mutually exclusive set of internal states (e.g., *D* or *not-D* for “detected” and “not detected”, respectively), and from the internal states overt responses are mapped. Thus, a primary difference between continuous- and discrete-state models is the number of internal

states that are posited. Discrete-state models assume a finite, and usually small, number of internal states, and continuous-state models assume a continuum of internal states (i.e.,  $-\infty < S < \infty$ ).

Discrete-state models are characterized by their state diagrams or transition matrices, which make the assumptions about how stimulus classes are mapped onto states (*state matrix*) and how states are mapped onto responses (*response matrix*) concrete. Here we will consider the double high-threshold model because it has been the subject of several recent empirical tests (e.g., Qin et al., 2001; Slotnick et al., 2000; Yonelinas, 1997, 1999).

The bottom panel of Figure 1 illustrates the double high-threshold model for yes–no tasks described by Macmillan and Creelman (1991). In this model, there are two types of stimuli,  $O$  and  $N$ , which are mapped onto three states.  $D_O$  and  $D_N$  are the states entered when  $O$  or  $N$  items are detected, and  $D_?$  is an indeterminate state (the state matrix is shown in Figure 1). The probability that an item will exceed a threshold and enter either  $D_O$  (for  $O$  items) or  $D_N$  (for  $N$  items) is  $q$ . With a probability  $(1 - q)$ , the item enters  $D_?$ . Hence, this model is known as a double high-threshold model, because unlike SDT, there is a threshold that only  $O$  items may exceed and a second threshold that only  $N$  items may exceed (Egan, 1958).

A response matrix for the yes–no task is shown in Figure 1. Here, items in either the  $D_O$  or  $D_N$  states are mapped with unity to the “O” and “N” responses, respectively. When items are in  $D_?$ , the classifier is assumed to guess “O” with probability  $p$  and “N” otherwise. The greater the value of  $p$ , the greater the bias to respond “O”.

### Constructing an ROC

There is an infinite number of hit rate (HR) and false-alarm rate (FAR) pairs that may be generated by a classifier for a given level of sensitivity, each depending on how bias he or she is to respond “yes” (because  $\beta$  and  $p$  are continuous variables, for example). HRs and FARs increase as the tendency to respond “yes” increases. An ROC plots FAR–HR pairs obtained under the assumption that bias varies but sensitivity is held constant. Discrete- and continuous-state models often make different predictions about the form of the ROC (e.g., linear vs. nonlinear), and for this reason it has been the subject of hypothesis testing (Qin et al., 2001; Slotnick et al., 2000; Yonelinas, 1997, 1999). The predictions of each model will be considered in detail in the next section, but to derive the proper predictions, it is important to first consider the manner in which the ROC is created. Here we briefly consider the yes–no and ratings methods.

In order to construct an ROC, response bias or guessing must be manipulated and a FAR–HR pair must be collected for each level of bias. Several methods for manipulating the criterion make use of the yes–no task, where the classifier is to determine whether  $t$  is a member of  $O$  by responding either “yes” or “no” (see Green & Swets, 1966). Another method to construct an ROC is based on subjective ratings (e.g., Egan, 1958). The classifier’s task in a ratings experiment is to assign  $t$  to one of  $n$  partitions of the classification scale, each reflecting a degree of confidence that  $t$  is in class  $O$ . For example, an item may be determined to be “old” or “new,” and the subject indicates whether he or she has either a low, moderate, or high level of confidence associated with that response. Thus, for this example, there are six levels of confidence

that an item is “old,” with an “old-high” being the highest and a “new-high” being the lowest.

The important thing to note about constructing ROCs is that the yes–no and ratings tasks are different. The yes–no task is a binary choice, and both the SDT and double high-threshold models are, at least in form, suitable for describing this behavior. However, the ratings task involves a choice between more than two responses. Therefore, yes–no models must be altered in order to accommodate the ratings task.

In the next sections, the yes–no and ratings SDT models are shown to predict the same ROC regardless of the method used for construction (Green & Swets, 1966). In contrast, the yes–no and ratings double high-threshold models can make different predictions about the form of the ROC depending on the assumptions concerning how states are mapped to responses.

### The SDT ROC

For the yes–no task, SDT assumes that the subject adopts a different level of  $\beta$  for each level of the biasing variable manipulated by the experimenter. Conditions produce less stringent criteria (i.e., subjects are more likely to respond “O”) when the subject perceives an “O” response as being more favorable or the probability of encountering an  $O$  item as more likely (Green & Swets, 1966). An equal-variance SDT ROC is shown in Figure 2, which is nonlinear (Green & Swets, 1966) and regular. For a regular ROC, there exists for every  $HR = y$ , a corresponding  $FAR = x$ , where  $0 < y < 1$  and  $0 < x < 1$  (Birdsall, 1966; Swets, 1986).

For the ratings task, SDT assumes the classifier maintains an ordered set of criteria (e.g.,  $c = c_1, c_2, \dots, c_n$ ) associated with different levels of confidence. At test, the subject assigns one level of confidence to each test item. A confidence rating,  $r_K$ , is made for  $t$  when  $S$  exceeds the  $K$ th level of confidence,  $c_K$ , but not the more strict  $c_{K-1}$ . It is further assumed that if  $S$  exceeds  $c_K$  then it exceeds  $c_{K+1}$ . Given  $n$  different levels of confidence,  $n - 1$  interdependent points are obtained by plotting the percentage of  $O$  items against the percentage of  $N$  items that were associated with a level of confidence greater than or equal to  $c_K$ .

The only difference between the yes–no and ratings tasks is the number of criteria with which  $S$  is compared. The assignment of responses to  $S$  is made in the same manner for both the yes–no and ratings tasks. Thus, the SDT ratings ROC is identical to the SDT yes–no ROC (Green & Swets, 1966; Larkin, 1965).

### Double High-Threshold ROC

Discrete-state models assume that different points on the ROC are associated with different degrees of bias,  $p$ , to guess “O”. The double high-threshold yes–no ROC, shown in the bottom panel of Figure 2, is nonregular because it intercepts the boundaries of ROC space at two points,  $[q, 0.0]$  and  $[1.0, (1 - q)]$ . These points are related by a linear function (Macmillan & Creelman, 1990, 1991). The linear increase is the result of adding the same ratio of hits and false alarms for each increment of  $p$ . The response matrix in Figure 1 shows that even when  $p = 0$ , the subject will sometimes respond “O” because a proportion,  $q$ , of  $O$  items is not subject to guessing (because items in the  $D_O$  state are always called “O”).

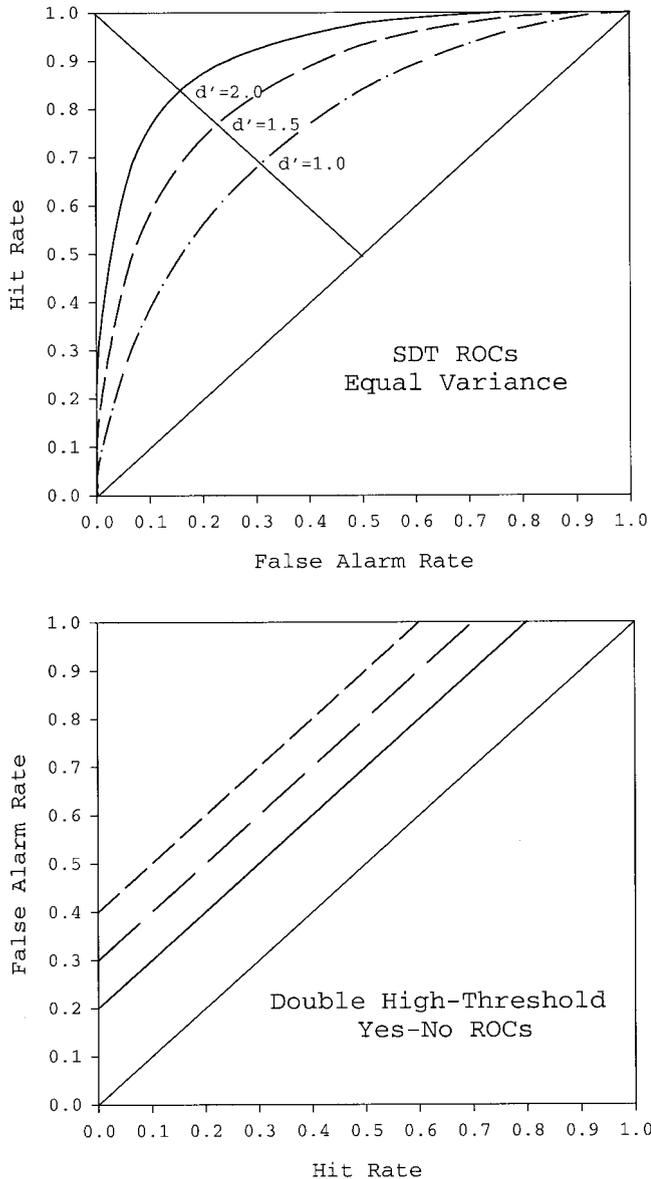


Figure 2. Equal-variance, signal detection, yes-no, and ratings receiver operating characteristics (ROCs), and double high-threshold, yes-no ROCs. SDT = signal detection theory.

Likewise when  $p = 1.0$ , the subject will respond "N" with probability  $q$ .

The double high-threshold ratings ROC is not predicted as easily. The primary difficulty is in specifying the response matrix (i.e., how states are mapped to responses). Nachmias and Steinman (1963) point out that a number of different response matrices are possible for the ratings task. Larkin (1965), Broadbent (1966), and Krantz (1969) show that only under certain assumptions does a low-threshold model predict a bilinear ratings ROC. Under different assumptions, the low-threshold model predicts a nonlinear ROC that is indistinguishable from the SDT ROC.

Here, the double high-threshold model is shown to be able to predict both linear and nonlinear ratings ROCs. First, one should

consider the response matrix in Table 1, which predicts linear ratings ROCs, given the state matrix in Figure 1. In this model, items entering the  $D_O$  or  $D_N$  states are always assigned the highest and lowest degrees of confidence, respectively, but items entering the  $D_I$  state are never assigned the most extreme confidence ratings. Rather, items in  $D_I$  are randomly assigned an intermediate confidence rating, each with probability  $p_K = 1/(n - 2)$ . The ROC for this ratings model is shown in the bottom panel of Figure 2. It intercepts the boundaries of ROC space at  $[q, 0.0]$  and  $[1.0, (1 - q)]$ , and these points are connected by a linear function. The resulting ROC is linear because the same proportion of  $N$  and  $O$  items is assigned to each intermediate confidence rating. This ROC was generated on the basis of the assumption that intermediate ratings are assigned randomly to items in  $D_I$ . It can be shown, however, that this assumption can be relaxed to include any response matrix, as long as the items entering the  $D_O$  and  $D_N$  states are always assigned the highest and lowest degrees of confidence, respectively.

Next, one should consider a double high-threshold model that predicts a nonlinear ratings ROC and whose response matrix is illustrated in Table 2. In this model, items in the  $D_O$  and  $D_N$  states may be mapped to any one of  $n$  confidence ratings. Specifically, it is assumed that items in  $D_O$  are mapped to  $c_K$ , with probability  $w_K$  (where  $0 < w_K < 1$ ), and items in  $D_N$  are mapped to  $c_K$ , with probability  $y_K$  (where  $0 < y_K < 1.0$ ).  $w_K$  and  $y_K$  are negatively related such that items detected as  $O$  are more likely to be assigned a relatively high confidence rating, and items detected as  $D_N$  are more likely to be assigned a relatively low confidence rating. That is,  $w_K/y_K$  is a negative function of  $K$  (where 1 is the highest confidence rating, and  $n$  is the lowest confidence rating). Items in  $D_I$  are randomly assigned a confidence rating including the most extreme ones with equal probability,  $p_K = 1/n$ . Thus,

$$P(r_K|O) = w_K q + b \quad (3)$$

and

$$P(r_K|N) = y_K q + b, \quad (4)$$

where  $b = p_K(1 - q)$ .

The ROC for this ratings model is illustrated in Figure 3, and its response matrix is listed in Table 2. The ratings ROC is nonlinear because different proportions of  $O$  and  $N$  items are being assigned to  $c_K$  depending on the ratio  $w_K/y_K$ . When  $w_K/y_K > 1.0$ , the slope of the ROC is greater than 1.0, and when  $w_K/y_K < 1.0$ , the slope of the ROC is less than 1.0. The ratings ROC is also regular,  $N$  items are sometimes assigned the highest confidence rating and  $O$  items are sometimes assigned the lowest confidence rating. Be-

Table 1  
Response Matrix for Double High-Threshold  
Ratings ROCs in Figure 2

State	Confidence ratings		
	1	$2 \leq K \leq (n - 1)$	$n$
$D_O$	1.0	0	0
$D_I$	0	$p_K$	0
$D_N$	0	0	1.0

Note. ROC = receiver operating characteristic.

Table 2  
Response Matrix for Double High-Threshold Ratings ROCs in Figure 3

State	Confidence ratings							
	1	2	3	4	5	6	7	8
D <sub>O</sub>	w <sub>1</sub> .600	w <sub>2</sub> .200	w <sub>3</sub> .080	w <sub>4</sub> .050	w <sub>5</sub> .030	w <sub>6</sub> .020	w <sub>7</sub> .010	w <sub>8</sub> .010
D <sub>?</sub>	p <sub>1</sub> .125	p <sub>2</sub> .125	p <sub>3</sub> .125	p <sub>4</sub> .125	p <sub>5</sub> .125	p <sub>6</sub> .125	p <sub>7</sub> .125	p <sub>8</sub> .125
D <sub>N</sub>	y <sub>1</sub> .010	y <sub>2</sub> .010	y <sub>3</sub> .020	y <sub>4</sub> .030	y <sub>5</sub> .050	y <sub>6</sub> .080	y <sub>7</sub> .200	y <sub>8</sub> .600

Note. ROC = receiver operating characteristic.

cause the ratings ROC is nonlinear and regular, it is empirically indistinguishable from the SDT ROC on the basis of form alone.

In this section, the double high-threshold model was shown to predict either a linear or nonlinear ratings ROC. A linear ratings ROC is predicted on the assumptions that items in D<sub>O</sub> and D<sub>N</sub> always receive the most extreme confidence ratings, and that items in D<sub>?</sub> are associated with an arbitrary distribution over the rating categories. A nonlinear ROC is predicted on the assumptions that D<sub>O</sub>, D<sub>?</sub>, and D<sub>N</sub> may be assigned any confidence rating, with  $w_K/y_K$  decreasing with  $K$  and  $p_K$  held constant.

This analysis shows that the critical factor that determines the shape of the double high-threshold ratings ROC is the response matrix. The models discussed here are meant to serve only as existence proofs. Other assumptions about how responses are mapped from states might predict similar nonlinear ratings ROCs. Hence, because SDT and the double high-threshold model predict a nonlinear ratings ROC, the form of the ratings ROC does not provide sufficient evidence to distinguish the two models. A linear ROC that is located substantially above chance would, however, be inconsistent with SDT.

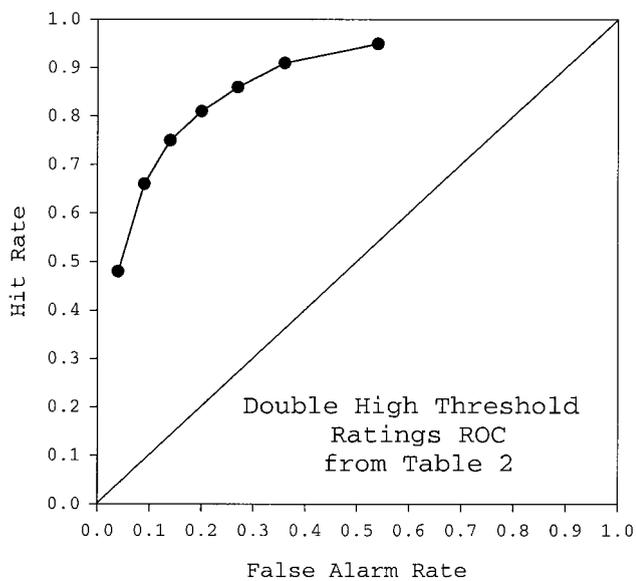


Figure 3. Nonlinear, double high-threshold confidence ratings receiver operating characteristic (ROC).

Symmetry of the ROC and Recognition Memory

Sometimes the symmetry of the recognition memory ratings ROC has been used to empirically test SDT and double high-threshold models (e.g., Slotnick et al., 2000; Yonelinas, 1997, 1999). SDT predicts a nonlinear ROC that is symmetrical with respect to the minor diagonal when the variance of the  $O$  and  $N$  distributions are the same. The symmetry can be confirmed by plotting the ROC in  $z$ -transformed ROC space because a unit-slope linear  $z$ -ROC is predicted when  $\sigma_N = \sigma_O$  (Green & Swets, 1966). When  $\sigma_N = \sigma_O$  is not assumed and the variance of the  $O$  distribution is greater than the variance of the  $N$  distributions, the SDT ROC is predicted to be asymmetrical with respect to the minor diagonal and the slope of the  $z$ -ROC is predicted to be less than 1.0 (Green & Swets, 1966). In fact, for recognition memory, asymmetrical ROCs and linear  $z$ -ROCs with a slope of about .75 are usually observed (e.g., Egan, 1958; Ratcliff & McKoon, 1991).

A double high-threshold model predicts asymmetrical ratings ROCs when the rate of change of  $w_K$  with respect to  $K$  is greater than the rate of change of  $y_K$  with respect to  $K$ , because the slope of the ROC equals  $w_K/y_K$ . One such response matrix is defined in Table 3, and it assumes that  $w_K$  decreases with  $K$  and  $y_K$  remains constant (i.e., confidence ratings are assigned randomly to items in D<sub>N</sub>). The predicted ratings ROC and the same one plotted in  $z$ -space are shown in the top and bottom panels of Figure 4, respectively. The ratings ROC is nonlinear, and the ratings  $z$ -ROC is linear (Slope = .74, as shown in Figure 4).<sup>1</sup> This is similar to what is typically observed (e.g., Egan, 1958; Ratcliff & McKoon, 1991).

In this section, it was shown that the unequal-variance SDT model and the double high-threshold model predict the empirical recognition ratings ROC. In addition, Macmillan and Creelman (1991) showed that a continuous-state model can predict a linear ratings ROC. Thus, neither the nonlinearity nor the asymmetry of the empirical ratings ROC is likely to be informative in distinguishing between continuous- and discrete-state models.

<sup>1</sup> Whereas the quadratic component of a linear regression on the ROC in Figure 4 increases the amount of variance accounted for (linear  $R^2 = .92$ ; quadratic  $R^2 = .99$ ), the quadratic component of a linear regression on the  $z$ -ROC adds almost nothing to the accounted-for variance (linear  $R^2 = .99$ ; quadratic  $R^2 = .99$ ).

Table 3  
Response Matrix for Double High-Threshold Ratings ROCs in Figure 4

State	Confidence ratings							
	1	2	3	4	5	6	7	8
$D_O$	$w_1$ .600	$w_2$ .200	$w_3$ .080	$w_4$ .050	$w_5$ .030	$w_6$ .020	$w_7$ .010	$w_8$ .010
$D_T$	$p_1$ .125	$p_2$ .125	$p_3$ .125	$p_4$ .125	$p_5$ .125	$p_6$ .125	$p_7$ .125	$p_8$ .125
$D_N$	$y_1$ .125	$y_2$ .125	$y_3$ .125	$y_4$ .125	$y_5$ .125	$y_6$ .125	$y_7$ .125	$y_8$ .125

Note. ROC = receiver operating characteristic.

### Symmetry of the ROC and Source Memory

In a typical yes–no source-memory experiment, subjects are presented stimuli from two different sources to remember. At test,

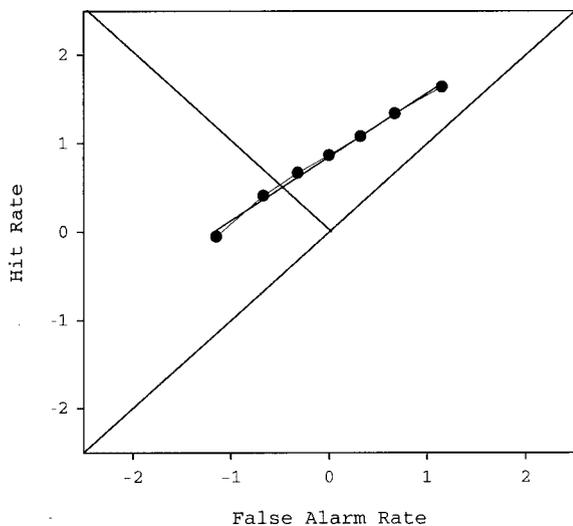
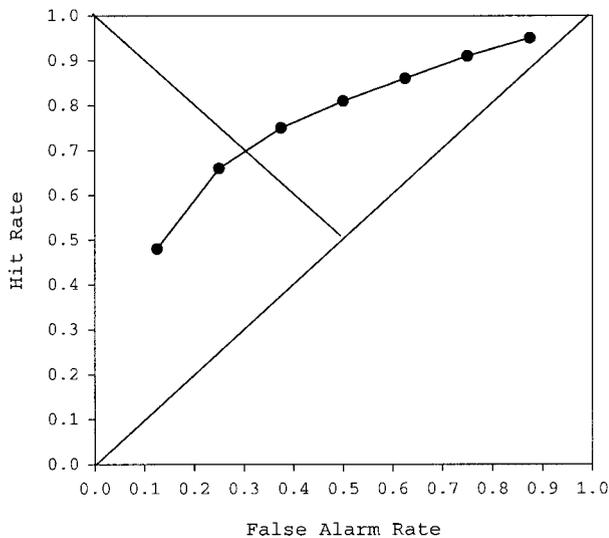


Figure 4. Asymmetrical, double high-threshold confidence ratings receiver operating characteristic (ROC), and z-ROC.

$O$  and  $N$  stimuli are presented, and the subject's tasks are to first determine if a stimulus was presented by either source (i.e., yes–no recognition), and if so, to determine which source presented it (i.e., a forced choice between the sources). Yonelinas (1999) used a ratings method to construct ratings ROCs for both the recognition and source memory parts of the task. In the ratings source-memory experiment, the subject's task is to determine what level of confidence he or she has that a stimulus is  $O$ , and to determine at what level of confidence he or she can identify its source. Yonelinas (1999) found that the recognition ratings ROC was nonlinear and asymmetrical, but that the source-memory ratings ROC was linear.

Yonelinas (1999) concluded that the recognition performance is well described by a dual-process model combining an equal-variance SDT model with a double high-threshold model (Yonelinas, 1999) because the recognition ratings ROC was nonlinear. He further concluded that the source memory performance is well described by only the high-threshold component of his dual-process model because the source-memory ratings ROC was linear (Yonelinas, 1999). However, the double high-threshold model can predict both linear and nonlinear/asymmetrical ratings ROCs (as was shown above). Thus, these findings provide little support for the dual-process assumption of the model.

The double high-threshold model makes different predictions about the form of a ratings ROC depending on the response matrix that is assumed. Thus, one way to predict a nonlinear recognition ROC and linear source memory ROC is if the subject uses different response strategies for the two tasks. Table 4 lists the percentages of ratings assigned to different types of stimuli collapsed over Yonelinas's (1999) experiments. The collapsed data are used to plot the composite recognition and source-memory ratings ROCs for these experiments, shown in Figure 5. Overall, the ratings ROCs in Figure 5 are qualitatively similar to the ratings ROCs from the individual experiments (Yonelinas, 1999). The recognition ratings ROC is asymmetrical and nonlinear, but the source memory ratings ROC is more nearly linear. There is a slight bend in the source-memory ratings ROC, but it appears more linear than the recognition rating ROC.

The distribution of confidence ratings in Table 4 can be used to work backward using Equations 3 and 4 to estimate the  $w_K$  and  $y_K$  for each level of confidence (where  $p = .167$  and  $q = .750$ ). If different response strategies are used for recognition and source memory (e.g., Riefer, Hu, & Batchelder, 1994), these will be observed in the values of  $w_K$  and  $y_K$ .

Several trends are apparent in the parameter estimates reported in Table 4 (along with  $w_K/y_K$ ). First, the distributions of  $w_K$  and  $y_K$

for recognition are highly skewed such that the most extreme ratings are used most often. In addition,  $w_K$  is more highly skewed than  $y_K$ . This is consistent with the double high-threshold model described earlier, in which  $w_K$  changes more rapidly than  $y_K$  (see Table 3).

For the source memory task, subjects tended to use more of the ratings scale than they used for recognition. The distributions tend to be slightly bimodal with one mode appearing at one extreme rating and the other mode appearing at an intermediate rating. The greater similarity of the source-memory distributions, relative to the recognition data, makes the  $w_K/y_K$  ratios change by smaller amounts over  $K$ , and thus the source-memory ROC is flatter.

Conclusions

A single-process double high-threshold model accounts for the Yonelinas's (1999) source memory and recognition findings by assuming that the subject adopts a different response matrix for each task. Erdfelder and Buchner (1998) used similar reasoning to show that their discrete-state model can account for the yes-no and ratings ROCs for an inclusion-exclusion recognition procedure. Thus, the forms of source and recognition ratings ROCs cannot alone provide disconfirmation of a single-process double high-threshold model. Yonelinas (1999) showed that a dual-process model could also account for these findings. However, there exists no evidence from the data presented here that confirms a dual-process account and disconfirms a single-process account.

When considering the form of an ROC, what may distinguish continuous- and discrete-state yes-no models of detection does not necessarily distinguish continuous- and discrete-state ratings models of detection (Kintsch, 1967; Larkin, 1965; Lockhart & Mur-

Table 4  
Distribution of Confidence Ratings for Yonelinas (1999) and Double High-Threshold Model Parameter Estimates

Item type	Confidence ratings ( $C_K$ )					
	1	2	3	4	5	6
Recognition						
Targets						
$P(C_K)$	.49	.13	.11	.10	.10	.06
$w_K$	.60	.12	.09	.08	.08	.03
Lures						
$P(C_K)$	.08	.11	.14	.18	.23	.25
$y_K$	.05	.09	.13	.19	.25	.28
$w_K/y_K$	12.00	1.33	.69	.42	.32	.11
Source						
Targets						
$P(C_K)$	.35	.11	.16	.20	.12	.06
$w_K$	.41	.09	.17	.21	.10	.02
Lures						
$P(C_K)$	.11	.07	.16	.23	.16	.27
$y_K$	.09	.04	.16	.25	.16	.31
$w_K/y_K$	4.55	2.25	1.06	.84	.63	.06

Note.  $P(C_K)$  = probability of giving the  $K$ th confidence rating.  $w_K$  and  $y_K$  = probabilities of giving the  $K$ th confidence rating, given states  $D_O$  and  $D_N$ , respectively, and were derived from  $P(C_K)$ , assuming  $p = .167$  and  $q = .750$ .

Single Process Double-High Threshold Model Predictions for Source Memory and Recognition

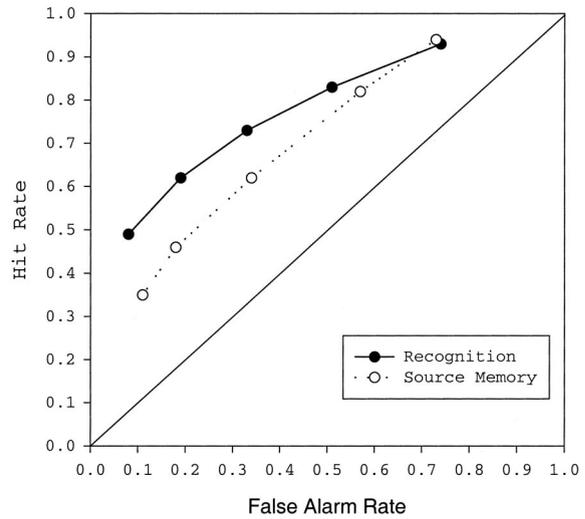


Figure 5. Recognition and source memory ratings receiver operating characteristics (derived from Yonelinas, 1999).

dock, 1970). A linear ROC is inconsistent with the single-process classical SDT account of source memory. However, performance is lower for source memory than for recognition, and the ratings ROC in Figure 5 is slightly bowed. Classical SDT predicts a flattening of the ROC with decreases in performance, as the ROC must converge on the main diagonal. In addition, Qin et al. (2001) and Slotnick et al. (2000) found nonlinear source-memory ratings ROCs. Thus, conclusions about whether the source-memory ratings ROC is consistent with SDT should, at this point, be made tentatively.

Because the inherent difficulty of drawing conclusions about the nature of detection from the form of ROCs derived from confidence ratings, yes-no procedures are sometimes considered more desirable than the ratings procedure when the form of the ROC is in question (Kintsch, 1967).<sup>2</sup> Relatively few yes-no recognition ROCs have been reported, but when they have been reported they have been nonlinear (e.g., Ratcliff, Sheu, & Gronlund, 1992). A nonlinear yes-no recognition ROC is inconsistent with threshold models of yes-no recognition. However, the form of the double high-threshold ratings ROC is determined by the response strategy, and therefore the form of the ratings ROCs is not diagnostic.

<sup>2</sup> The preference for yes-no ROCs is conditional on the assumption that the biasing manipulation does not also alter sensitivity. Especially for relatively complex tasks, like source memory, different strategies for remembering may be used that are dependent on the proportion of targets the subject expects to encounter or the amount of rewards or costs the subject expects on the basis of his or her performance.

References

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

- Batchelder, W. H., Riefer, D. M., & Hu, X. (1994). Measuring memory factors in source monitoring: Reply to Kinchla. *Psychological Review*, *101*, 172–176.
- Birdsall, T. G. (1966). The theory of signal detectability: ROC curves and their character. *Dissertation Abstracts International*, *28*, 1B. University of Michigan.
- Broadbent, D. E. (1966). Two-state threshold model and rating-scale experiments. *Journal of the Acoustical Society of America*, *40*, 244–245.
- Buchner, A., Erdfelder, E., & Vaterrodt-Plunnecke, B. (1995). Toward unbiased measurement of conscious and unconscious memory processes within the process dissociation framework. *Journal of Experimental Psychology: General*, *124*, 137–160.
- Egan, J. P. (1958). *Recognition memory and the operating characteristic* (No. AFCRC-TN-58-51). Bloomington, IN: Indiana University Hearing and Communications Laboratory.
- Erdfelder, E. & Buchner, A. (1998). Process-dissociation measurement models: Threshold theory or detection theory? *Journal of Experimental Psychology: General*, *127*, 83–96.
- Glanzer, M., Kim, K., Hilford, A., & Adams, J. K. (1999). Slope of the receiver operating characteristic in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 500–513.
- Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.
- Kinchla, R. A. (1994). Comments on Batchelder and Riefer's multinomial model for source monitoring. *Psychological Review*, *101*, 166–171.
- Kintsch, W. (1967). Memory and decision aspects of recognition learning. *Psychological Review*, *74*, 496–504.
- 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.
- Lockhart, R. S. & Murdock, B. B. (1970). Memory and the theory of signal detection. *Psychological Bulletin*, *74*, 100–109.
- Macmillan, N. A. & Creelman, C. D. (1990). Response bias: Characteristics of detection theory, threshold theory, and “nonparametric” indexes. *Psychological Bulletin*, *107*, 401–413.
- Macmillan, N. A. & Creelman, C. D. (1991). *Detection theory: A user's guide*. Cambridge, England: Cambridge University Press.
- 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.
- Qin, J., Raye, C. L., Johnson, M. K., & Mitchell, K. J. (2001). Source ROCs are (typically) curvilinear: Comment on Yonelinas (1999). *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*, 1110–1115.
- Ratcliff, R., & McKoon, G. (1991). Using ROC data and priming results to test global memory models. In W. E. Hockley & S. Lewandowsky (Eds.), *Relating theory and data; Essays in honor of Bennet B. Murdock* (pp. 279–296). Hillsdale, NJ: Erlbaum.
- Ratcliff, R., Sheu, C.-F., & Gronlund, S. D. (1992). Testing global memory models using ROC curves. *Psychological Review*, *99*, 518–535.
- Riefer, D. M., Hu, X., & Batchelder, W. H. (1994). Response strategies for source monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 680–693.
- Slotnick, S. D., Klein, S. A., Dodson, C. S., & Shimamura, A. P. (2000). An analysis of signal detection and threshold models of source memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 1499–1517.
- Swets, J. A. (1986). Form of empirical ROCs in discrimination and diagnostic tasks: Implications for theory and measurement of performance. *Psychological Bulletin*, *99*, 181–199.
- Yonelinas, A. P. (1997). Recognition memory ROCs for item and associative information: Evidence for a single-process signal-detection model. *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, and Cognition*, *25*, 1415–1434.
- Yonelinas, A. P. & Jacoby, L. L. (1996). Response bias and the process dissociation procedure. *Journal of Experimental Psychology: General*, *125*, 422–434.

Received January 17, 2001

Revision received August 9, 2001

Accepted August 9, 2001 ■