

Signal Detection Theory With Finite Mixture Distributions: Theoretical Developments With Applications to Recognition Memory

Lawrence T. DeCarlo
Teachers College, Columbia University

An extension of signal detection theory (SDT) that incorporates mixtures of the underlying distributions is presented. The mixtures can be motivated by the idea that a presentation of a signal shifts the location of an underlying distribution only if the observer is attending to the signal; otherwise, the distribution is not shifted or is only partially shifted. Thus, trials with a signal presentation consist of a mixture of 2 (or more) latent classes of trials. Mixture SDT provides a general theoretical framework that offers a new perspective on a number of findings. For example, mixture SDT offers an alternative to the unequal variance signal detection model; it can also account for nonlinear normal receiver operating characteristic curves, as found in recent research.

Signal detection theory (SDT) provides a theoretical framework that has been quite useful in psychology and other fields (see Gescheider, 1997; Macmillan & Creelman, 1991; Swets, 1996). A basic idea of SDT is that decisions about the presence or absence of an event are based on decision criteria and on perceptions of the event or nonevent, with the perceptions being represented by probability distributions on an underlying continuum. Thus, in its simplest form, the theory considers two basic aspects of detection—the underlying representations, which are interpreted as psychological distributions of some sort (e.g., of perception or familiarity), and a decision aspect, which involves the use of decision criteria to arrive at a response.

The present article extends SDT by viewing detection as consisting of an additional process. The result is a simple and psychologically meaningful extension of SDT that can be applied to any area of research where SDT has been applied. The approach is illustrated with applications to research on recognition memory, where the additional process can be interpreted as attention. In particular, the basic idea is that presentation of a signal shifts the location of the underlying distribution only if the observer is attending to the signal; otherwise, the distribution is not shifted or is only partially shifted. As a result, trials with a signal presentation consist of a mixture of two (or more) latent classes of trials, which can be interpreted as being attended and nonattended trials (or as two different levels of processing of the stimuli). Apart from that, the theory is the same as in conventional SDT, in that the underlying representations are used together with response criteria to arrive at an observed response. I show, however, that this simple extension of SDT is quite powerful and can account for a variety of findings across several areas of research. For example, the mixture approach offers an alternative to the unequal variance

signal detection model (Green & Swets, 1966), which has been the standard model for many years, and provides a different interpretation of normal receiver operating characteristic (ROC) curves with slopes less than unity; it can also account for nonlinear normal ROC curves, as found in recent research. I also show that mixture SDT offers new insights into the data and suggests new directions for future research; mixture SDT recognizes the influence of an additional process in detection and greatly enlarges the scope of classical SDT.

SDT With Finite Mixture Distributions

Consider the basic signal detection experiment where, on each trial, either a signal or noise is presented. For example, in recognition memory research, the signal might consist of words that were shown on a list studied earlier (old items), and noise might consist of words that were not on the list (new items). In SDT, the effect of a presentation of a new item is represented by a probability distribution on an underlying continuum, and the effect of a presentation of an old item is represented by a probability distribution with a different location; the distance between the two distributions reflects the observer's ability to detect. In the extended version of SDT, the effect of a presentation of an old item is represented by two (or more) probability distributions rather than one. The two distributions correspond to different representations of the signal and can be motivated in more than one way, depending on the particular research context. A general interpretation is that the distributions arise from different levels of attention (or processing) to the signal (during the study period for memory research or during the test period for psychophysical research). It should be noted that although the focus here is on two levels (which appears to be adequate for the data analyzed below), the mixture approach is general and can include more than two levels (i.e., latent classes) if necessary; a situation where more than two latent classes might be needed (repetition) is noted below.

Figure 1 illustrates the basic ideas. The distribution on the left (with mode Ψ_N) represents the familiarity of new items, as in the usual version of SDT as applied to recognition memory. For old items, on the other hand, there are two distributions instead of one.

LEM and Mplus programs for some of the examples discussed here are available at <http://www.columbia.edu/~ld208>.

Correspondence concerning this article should be addressed to Lawrence T. DeCarlo, Department of Human Development, Teachers College, Columbia University, Box 118, 525 West 120th Street, New York, New York 10027. E-mail: decarlo@exchange.tc.columbia.edu

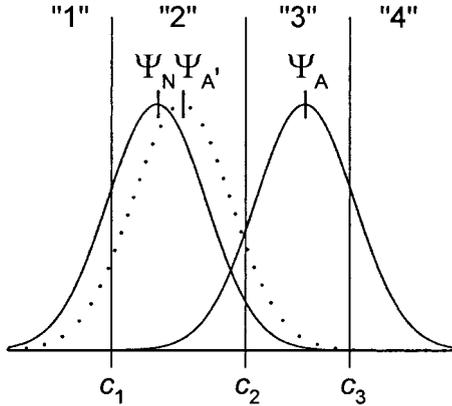


Figure 1. An example of signal detection theory with mixture distributions. Ψ_N , Ψ_A , and $\Psi_{A'}$ are the modes of the reference (noise or new), A, and A' distributions, respectively; c_1 , c_2 , and c_3 are three response criteria that delineate responses of 1, 2, 3, and 4.

One distribution, denoted as the A distribution (with mode Ψ_A), is associated with items that were attended to during the study period, and so the distribution is shifted to the right (higher familiarity), as in the usual situation. The other distribution, denoted as A' (represented by the dotted line), represents items that were processed at a different level, such as those not being attended to or only being partially attended to during the study period, and so their distribution has a different location than that for attended items. The result is that during a recognition test, responses on a proportion of the signal trials depend on the A distribution, and for the remaining proportion of signal trials depend on the A' distribution. If it was known which items received high and low attention, say, then classical SDT could be used to obtain estimates of the observer's ability to detect the A and A' items. However, the level of attention to each item is not known, so there are latent classes of items, with the result that observed responses to a signal arise from a mixture of the underlying distributions associated with each type of item.

A model for the response probabilities follows directly from the theory illustrated in Figure 1. First note that, for rating or binary responses, the equal variance normal theory signal detection model can be written as

$$p(Y \leq k|X) = \Phi[(c_k - d_A X)/\sigma],$$

where Y is the response variable (e.g., a confidence rating) with values of k (e.g., from 1 to 4), X is a dummy-coded variable (new item = 0, old item = 1), $p(Y \leq k|X)$ is the cumulative probability of a response of k or less given X , with $1 \leq k \leq K - 1$, where K is the number of response categories, Φ is the cumulative standard normal distribution function, c_k are response criteria (i.e., distances from the new item distribution), d_A is the distance of the A distribution from the new item distribution, and σ is the standard deviation of the underlying distributions, which can be set to unity without loss of generality (and so it is not included in the equations below). The model can also be written more generally for probability distributions other than the normal by replacing Φ with other cumulative distribution functions (through the use of inverse link functions; see DeCarlo, 1998).

Mixture SDT extends the basic signal detection model by assuming that there are different latent classes of old items, associated with which are different underlying distributions, as illustrated in Figure 1. With the assumption that the latent classes are mutually exclusive and exhaustive, a mixture model for SDT is

$$p(Y \leq k|X) = \lambda \Phi(c_k - d_{A'} X) + (1 - \lambda) \Phi(c_k - d_A X), \quad (1)$$

where $d_{A'}$ is the distance of the A' distribution from the new item distribution and λ is the mixing proportion (with $0 \leq \lambda \leq 1$), interpreted here as indicating the proportion of trials on which the signal was attended to. Equation 1 accounts for the observed responses through a mixture of signal detection models, with different detection parameters in each component of the mixture but the same response criteria; note that if $\lambda = 1$, then the model reduces to the equal variance normal theory signal detection model. Equation 1 can be further generalized by replacing Φ with other functions, which corresponds to mixtures of other distributions (like the logistic or extreme value), in which case the model is a mixture of generalized linear models (DeCarlo, 2000); an example is noted below.

A simplification that appears to apply in many situations, as shown below, is obtained with $d_{A'} = 0$, in which case the model is

$$p(Y \leq k|X) = \lambda \Phi(c_k - d_A X) + (1 - \lambda) \Phi(c_k). \quad (2)$$

Equation 2 has a simple yet appealing interpretation: The restriction $d_{A'} = 0$ implies that rather than some items being only partially attended to during the study period, they are not attended to at all, so they are equivalent to new items during the subsequent test (and the nonattended A' distribution has the same location as the new item distribution). Thus, in its simplest form, the theory views responses on signal trials as being determined by a mixture of signal and noise distributions, corresponding to attended and nonattended trials, for example.

Equation 2 generalizes the equal variance normal signal detection model with the introduction of one parameter, the mixing proportion λ . The mixture model provides a basic way to generalize the simple signal detection model that differs from the traditional generalization, the unequal variance signal detection model (Green & Swets, 1966), which allows the variances of the underlying distributions to differ. For normal (Gaussian) distributions, the unequal variance SDT model can be written as

$$p(Y \leq k|X) = \Phi[(c_k - d_N X)/\sigma_s^X], \quad (3)$$

which generalizes the equal variance normal model by introducing the parameter $\sigma_s > 0$; d_N is the detection parameter scaled with respect to the standard deviation of the noise distribution (see DeCarlo, 1998). With the assumption that $\sigma_n = 1$, $1/\sigma_s$ gives the ratio of the standard deviations of the new and signal distributions, which corresponds to the slope of the ROC curve on inverse normal coordinates. With respect to the unequal variance normal signal detection model, Green and Swets (1966) noted that

The justification for the Gaussian model with unequal variance is, we believe, not to be made on theoretical but rather on practical grounds. It is a simple, convenient way to summarize the empirical data with the addition of a single parameter. (p. 79)

Here it is shown that mixture SDT provides an alternative extension of equal variance SDT that, like the unequal variance model, offers a simple summary of the data with the addition of a single parameter (i.e., Equation 2). In addition, an attractive aspect of the mixture approach is that it can easily be theoretically motivated.

It should also be noted that the unequal variance extension of SDT raises a somewhat neglected complication. In particular, with respect to introducing a parameter to allow the variance to vary, Green and Swets (1966) noted

Not only does this ad hoc assumption introduce another free parameter, but it creates a set of rather knotty theoretical problems. If the ratios of the standard deviations of the noise and signal distributions are unequal, then, as discussed above, the decision axis x is no longer monotonic with likelihood ratio. (p. 79)

Figure 2 illustrates the problem. The upper panel of the figure shows two normal distributions with different variances (noise distribution, $M = 0.0, SD = 1.0$; signal distribution, $M = 0.5, SD = 3.0$) and two possible locations of a response criterion, indicated by lines A and B. Note that if an observer located a yes-no criterion at A (i.e., they reported a signal above that value and noise below that value), then they would be reporting a signal when the probability of a signal was indeed higher than that of noise and so their hit rate would be greater than their false alarm rate, as can be seen in the figure (i.e., the area under the signal curve to the right of the A criterion is clearly larger than that for

noise). This is also shown by the lower panel, which presents the linearized (i.e., inverse normal or z -transformed) ROC curve that corresponds to the distributions shown in the upper panel and the locations of A and B. If, on the other hand, an observer located their criterion at Point B, then they would be reporting a signal when in fact the probability of noise was higher (note in this case that the area under the signal curve to the left of the B criterion is clearly larger than that for noise, so the area to the right is smaller for signal), and so their hit rate would be lower than their false alarm rate, as shown by Point B in the lower panel; they would in essence perform worse than if they were simply guessing. This reflects the lack of monotonicity (in the likelihood ratio) that Green and Swets (1966) referred to as a “knotty theoretical problem” (p. 79). By simply changing the location of their response criterion, an observer would appear to go from good to below chance performance, even though there was actually no change in their detection, that is, the distance between the distributions was constant and greater than zero.

It is important to recognize that the above complication only arises for the unequal variance extension of SDT and not for the mixture model of Equation 2. The structure of the models (Equations 2 and 3) also reflects the difference: The mixture model is additive within the latent classes, because it is assumed that the effect of an event is to simply shift the location of the underlying distribution, with the variance remaining constant, whereas the unequal variance model is nonlinear in the parameters; it assumes that an event has two effects, in that there is both a shift in location and a change in the variance of the underlying distribution.

Inverse Normal ROC Curves for Mixture Distributions

A signal detection model implies a set of ROC curves. For the mixture SDT model with normal underlying distributions, the ROC curves within each component of the mixture have unit slopes on inverse normal coordinates. The ROC curves for the mixture, however, do not necessarily have unit slopes and are more complex. The curves that follow from the theory can be derived by noting that a normal ROC plot is simply a plot of the inverse normal transform of $p(Y > k | X = 0)$ on the abscissa against the inverse normal transform of $p(Y > k | X = 1)$ on the ordinate. For example, for the mixture model of Equation 2, it follows that the normal ROC curve is a plot of

$$\Phi^{-1}[1 - \Phi(c_k)],$$

on the X axis (i.e., the inverse normal transform of one minus the equation for false alarms, that is, Equation 2 with $X = 0$) against

$$\Phi^{-1}\{1 - [\lambda\Phi(c_k - d_A) + (1 - \lambda)\Phi(c_k)]\}$$

on the Y axis (i.e., the inverse normal transform of one minus the equation for hits). The above are simply the equations for the theoretical mixture ROC curves on inverse normal coordinates. Note that the fitted ROC curves presented below are obtained by substituting the maximum likelihood estimates of d_A and λ in the above and plotting the curves generated by varying c_k .

Figure 3 presents ROC curves for various mixing proportions λ and values of d_A (for Equation 2). The upper left panel shows an example where λ varies and d_A remains constant, and it is apparent that as λ decreases the curves are pulled down toward the diagonal,

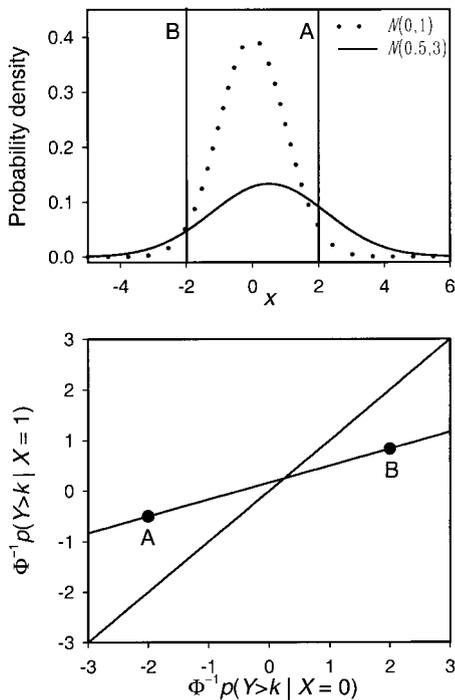


Figure 2. The upper panel shows two unequal variance normal distributions and two locations of a response criterion, indicated by vertical lines denoted as A and B. For the left distribution, $M = 0.0, SD = 1$; for the right distribution, $M = 0.5, SD = 3$. The lower panel shows the corresponding receiver operating characteristic curves on inverse normal coordinates, with the locations of the criterion shown in the upper panel indicated as Points A and B.

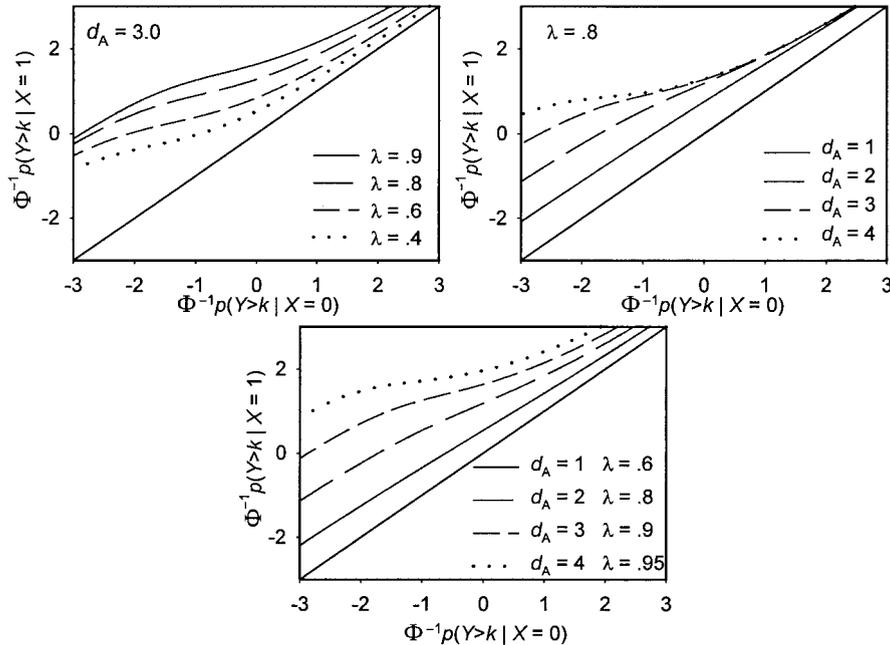


Figure 3. Examples of mixture receiver operating characteristic curves on inverse normal coordinates.

so that from the conventional perspective detection appears to be lower; the effect is also larger for the right side of the curves, thus giving them curvature and slopes less than unity. The upper right panel shows that, for constant λ , the curves are near linear for small d_A , but the curvature increases and the slope decreases as d_A increases. The lower panel shows an example where both d_A and λ change; from the conventional perspective, the curves are consistent with a decrease in the slope of the normal ROC curve as detection increases (a result that is commonly found). Overall, the figure shows that, with the addition of one parameter to the model (λ), a variety of ROC curves are generated. Examples of ROC curves consistent with the patterns shown in Figure 3 are readily found in the memory literature; for example, in addition to those presented below, see empirical ROC curves presented by Glanzer, Kim, Hilford, and Adams (1999); Ratcliff, McKoon, and Tindall (1994); and Yonelinas (1994, 1999). It should be kept in mind that, although the mixture ROC curves shown in Figure 3 are nonlinear, the unmixed ROC curves are simply linear with unit slopes (on inverse normal coordinates).

Because data obtained in signal detection experiments usually fall within the range of -2 to 2 (i.e., the false alarm and hit rates fall within the range of .02 and .98), and because the mixture ROC curves are near linear within this range and have slopes less than unity, as shown by Figure 3, data previously found to be well described by the unequal variance normal model will tend to be well described by mixture SDT, as shown in the next section. The interpretation of the results differs, however, and this is the aspect that is stressed here. Also considered below are situations where normal ROC curves show departures from linearity, which are not consistent with unequal variance normal theory, whereas the departures are well described by mixture SDT.

The mixture SDT model can be viewed as a hybrid of a structural equation model and a latent class model (DeCarlo, 2000), so

it brings with it a well-developed statistical methodology used, for example, in latent class analysis (see Dayton, 1998) and for mixture models (e.g., McLachlan & Peel, 2000); some aspects of fitting and testing the models are illustrated here. The Appendix discusses some additional details, such as how the model can be fit using the software LEM (Vermunt, 1997), which is freely available on the Internet (the Web site is given in the Appendix); the mixture model (with logistic distributions) can also be fit using Version 2 of the structural equation software Mplus (Muthén & Muthén, 1998). Some LEM and Mplus programs for the data discussed below are available at my Web site (<http://www.columbia.edu/~ld208>).

Some Applications of Mixture SDT: Recognition Memory

Recent research has focused on the use of ROC curves to evaluate different theories of recognition memory (e.g., Glanzer, Kim, et al., 1999; Ratcliff et al., 1994; Ratcliff, Sheu, & Gronlund, 1992). One aspect of this research has been an examination of the effects of different variables on the slope of the ROC curve. The results have been interpreted in terms of effects of the variables on the variances of the underlying distributions, and this in turn has influenced theories of recognition memory. It is shown here that, from the perspective of mixture SDT, the results can be interpreted in terms of effects on the mixing proportion, with the variance remaining constant.

Several examples are used to illustrate the application and implications of the mixture extension of SDT. The focus is on experiments where variables that might affect attention were manipulated and for which published data are readily available. The first example considers the effect of brief presentation times. The use of brief presentation times is of interest in light of mixture SDT because it seems reasonable to expect that more items will be

missed (not attended to) when presentation times are brief. Thus, brief presentation times should affect the parameter λ , if indeed it reflects attention. This prediction is examined using published data from recognition memory experiments of Ratcliff et al. (1994); the results for the mixture model are compared and contrasted with those for the unequal variance model. A second example examines experiments where both word frequency and presentation time were manipulated. The effects of word frequency have been considered in several theories, such as attention likelihood theory (Glanzer, Adams, Iverson, & Kim, 1993), where it is assumed that lower frequency words receive more attention than higher frequency words. The effect of word frequency on λ is examined and results for the mixture model are again compared with those for the unequal variance model.

Presentation time. Experiments 1 and 2 of Ratcliff et al. (1994) used brief presentation times during the study period: 50 ms, 100 ms, 200 ms, and 400 ms. Mixed and pure conditions were also used: For the pure condition, all the words were presented for the same amount of time during the study period; for the mixed condition, different presentation times were mixed together during the study period. The results for the pure conditions are presented here (the results for the mixed conditions are virtually identical, as noted by Ratcliff et al., 1994, and as is evident from inspection of their ROC curves). During the test phase, participants indicated how sure they were that a word was old or new by using a six-category ordered response scale, coded here as 1 = *sure new*, 2 = *probably new*, 3 = *maybe new*, 4 = *maybe old*, 5 = *probably old*, 6 = *sure old*.

Figure 4 shows the data on inverse normal coordinates, along with ROC curves for the fitted mixture model (Equation 2). In all cases, the mixture ROC curves closely describe the data; note that the fitted lines are slightly curved but are near linear. The figure shows that decreasing presentation time systematically decreased detection, in that the curves are closer to the diagonal.

Table 1 presents results for fits of the equal variance normal SDT model, the mixture SDT model (Equation 2), and the unequal variance SDT model (Equation 3). The table shows that the likelihood ratio (*LR*) goodness-of-fit statistics¹ are considerably

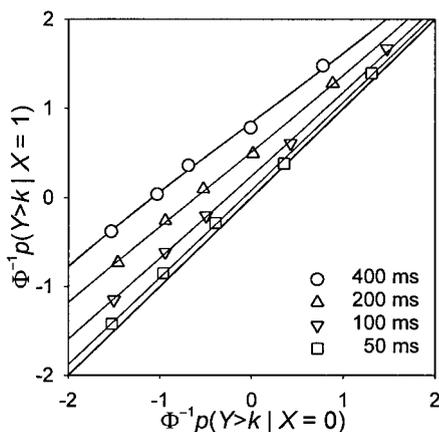


Figure 4. Data and fitted mixture receiver operating characteristic curves on inverse normal coordinates for Experiments 1 and 2 of Ratcliff, McKoon, and Tindall (1994). The different presentation times are indicated in the figure legend.

Table 1
Goodness-of-Fit Statistics and Information Criteria for Experiments 1 and 2 of Ratcliff, McKoon, and Tindall (1994)

Model	LR	df	p	AIC	BIC
400-ms study time (N = 4,184)					
Mixture SDT	15.77	3	< .01	13,896	13,940
Unequal variance SDT	17.79	3	< .01	13,898	13,942
Equal variance SDT	72.05	4	< .01	13,950	13,988
200-ms study time (N = 4,451)					
Mixture SDT	4.90	3	.18	15,297	15,342
Unequal variance SDT	5.75	3	.12	15,298	15,343
Equal variance SDT	39.54	4	< .01	15,330	15,368
100-ms study time (N = 4,176)					
Mixture SDT	2.37	3	.50	13,603	13,648
Unequal variance SDT	2.97	3	.40	13,604	13,648
Equal variance SDT	10.12	4	.04	13,609	13,647
50-ms study time (N = 4,295)					
Mixture SDT	5.51	3	.14	14,315	14,359
Unequal variance SDT	5.51	3	.14	14,315	14,359
Equal variance SDT	6.24	4	.18	14,313	14,351

Note. LR = likelihood ratio goodness-of-fit test; AIC = Akaike's information criterion; BIC = Bayesian information criterion; SDT = signal detection theory; mixture SDT = Equation 2 in text; unequal variance SDT = Equation 3 in text.

smaller for the mixture and unequal variance models than for the equal variance model in every case (significant values of the goodness-of-fit statistic indicate rejection of the null hypothesis that the model fits), except for the 50 ms condition, where the equal-variance model appears to be adequate. Note that, although the *LR* statistics for the 400 ms condition are significant, Figure 4 shows that the fitted mixture ROC curve closely describes the data; the significant statistics in this case reflect in part the large sample size ($N = 4,184$), which gives high power for the tests. This shows why, when assessing the adequacy of a model, visual inspection of the fitted ROC curves should be used in addition to formal tests. In spite of the high power, note that, for the mixture model, tests of $d_{A'} = 0$ did not lead to rejection of the null hypothesis in any case; an *LR* test can be performed by fitting Equations 1 and 2 and subtracting the $-2 \log$ -likelihoods, which gave, for 50 ms, $LR(1, N = 4,295) = 0.03, p = .86$; for 100 ms, $LR(1, N = 4,176) = 0.00, p = .98$; for 200 ms, $LR(1, N = 4,451) = 0.06, p = .80$; for 400 ms, $LR(1, N = 4,184) = 0.10, p = .75$. Thus, the simple one-parameter extension of Equation 2 (with $d_{A'} = 0$) appears to be adequate.

Also shown in Table 1 are information criteria (see Agresti, 1990; Burnham & Anderson, 1998), specifically Akaike's information criterion (AIC) and the Bayesian information criterion (BIC), which allow one to compare non-nested models; note that

¹ The chi-square goodness-of-fit statistics were also examined and were close in magnitude to the *LR* statistics in every case, so only the *LR* is reported.

the mixture and unequal variance signal detection models are not nested because one model cannot be obtained by restricting parameters of the other model. The information criteria incorporate a penalty for additional parameters, with smaller values of the criteria indicating a better model. Table 1 shows that the values of both the AIC and BIC for the mixture model are equal to or smaller than those for the unequal variance model in every case, although the difference tends to be small (and the simple equal variance model appears adequate for the 50-ms condition). Overall, the mixture model and unequal variance model appear to fit about equally well. Thus, Table 1 shows that what have typically been described as linear normal ROC curves with slopes less than unity are equally well described as normal mixture ROC curves. An important point to keep in mind is that, for mixture SDT, the variance of the signal distribution is equal to that of the noise distribution.

Table 2 presents maximum likelihood parameter estimates for the mixture model and for the unequal variance model. For the mixture model, detection decreases systematically with presentation time, and the estimated detection parameter is not significantly different than zero for the 50-ms presentation time (i.e., the estimate of d_A is less than its standard error). Thus, decreasing presentation time led to a decrease in detection.

Of particular interest is the effect of presentation time on the mixing proportion λ , interpreted here as a measure of attention. Table 2 shows that the estimates of λ decrease in magnitude as presentation time decreases (and the estimate is not significantly different than zero for the 50-ms presentation time). The interpretation of this result is simple yet compelling: The proportion of words attended to steadily decreased as presentation time decreased, with few if any words attended to for very brief presentation times.

The right side of Table 2 shows the parameter estimates obtained for the unequal variance model. With respect to detection, the results are similar to those for the mixture model, in that detection decreases as presentation time decreases. Note that the detection estimates are smaller in absolute magnitude than those found for mixture SDT; from the perspective of mixture SDT, this occurs because the mixing pulls the ROC curves down toward the diagonal, so that detection appears to be lower, as noted earlier in the discussion of the ROC curves shown in Figure 3. Table 2 also shows the estimates of the log of the standard deviation of the signal distribution (estimates of which are directly obtained because of the way the model is parameterized in LIMDEP; see the

Appendix) and the estimated ratio of standard deviations, obtained by exponentiating and taking the inverse of the estimate of $\log \sigma_s$ (using the natural log). The results show that the variance of the signal distribution decreases as presentation time decreases.

In summary, the results for presentation times are adequately described by both mixture SDT and unequal variance SDT. The interpretation of the results, however, differs considerably across the two approaches. From the perspective of mixture SDT, decreasing the presentation time led to a smaller proportion of items being attended to, with the variance of the underlying distributions remaining constant. The results provide important evidence that supports the interpretation of the mixing proportion as a measure of attention. From the perspective of unequal variance SDT, on the other hand, decreasing the presentation time led to a decrease in the variance of the signal distribution, and there is no obvious reason why the variance should be smaller; nevertheless, the interpretation in terms of the unequal variance model has had a considerable impact on the development of theories of recognition memory. Here I simply note that mixture SDT offers a simple and psychologically plausible interpretation of the results, in that the effect of presentation time on attention is as expected, whereas the interpretation in terms of the unequal variance model is not as clear; one could easily argue, for example, that the variance of the representation should increase as presentation time decreases (because there is less information about the signal), not decrease.

Word frequency. Effects of word frequency have been examined in a number of studies; the published data of Experiments 4 and 5 of Ratcliff et al. (1994), which have been analyzed in several articles (e.g., Glanzer et al., 1999; Ratcliff, Van Zandt, & McKoon, 1995), are analyzed here. Participants were presented with study lists consisting of either low- or high-frequency words, presented during the study phase for either 2 s or 5 s (Experiment 4) or for 1.5 s or 5 s (Experiment 5). During the test phase, participants responded on a six-category scale as to how sure they were that the word was old or new (with the same response labels as in Experiments 1 and 2, given above). The four pure-list conditions are analyzed here (i.e., the study lists consisted of only one type of word rather than the different types mixed together); the results are similar for the mixed lists, as shown by Ratcliff et al. (1994). For Experiment 4, the pooled data presented by Ratcliff et al. (1994) are analyzed; for Experiment 5, the individual data presented by Ratcliff et al. (1994) are analyzed.

Table 3 presents, for each of the four conditions in Experiment 4, results for fits of the equal variance normal SDT model,

Table 2
Parameter Estimates for Mixture Signal Detection Model for Experiments 1 and 2 of Ratcliff, McKoon, and Tindall (1994)

Study time	Mixture SDT		Unequal variance SDT		
	d_A	λ	d_N	$\log \sigma_s$	σ_n/σ_s
400 ms	1.49 (0.08)	0.69 (0.03)	1.07 (0.04)	0.23 (0.03)	0.79
200 ms	1.20 (0.11)	0.50 (0.05)	0.61 (0.04)	0.17 (0.03)	0.85
100 ms	0.87 (0.23)	0.31 (0.09)	0.26 (0.03)	0.07 (0.03)	0.93
50 ms	0.47 (0.92)	0.18 (0.35)	0.08 (0.03)	0.02 (0.03)	0.98

Note. Standard errors are in parentheses. SDT = signal detection theory; mixture SDT = Equation 2 in text; unequal variance SDT = Equation 3 in text.

Table 3
Goodness-of-Fit Statistics and Information Criteria for Experiment 4 of Ratcliff, McKoon, and Tindall (1994)

Model	LR	df	p	AIC	BIC
High freq 2 s (<i>N</i> = 2,816)					
Mixture SDT	3.98	3	.26	9,671	9,712
Unequal variance SDT	4.45	3	.22	9,671	9,713
Equal variance SDT	33.22	4	< .01	9,698	9,734
Low freq 2 s (<i>N</i> = 2,823)					
Mixture SDT	5.56	3	.14	9,000	9,042
Unequal variance SDT	5.19	3	.16	9,003	9,042
Equal variance SDT	52.47	4	< .01	9,045	9,081
High freq 5 s (<i>N</i> = 2,917)					
Mixture SDT	5.94	3	.11	9,758	9,800
Unequal variance SDT	6.36	3	.10	9,759	9,801
Equal variance SDT	28.32	4	< .01	9,779	9,815
Low freq 5 s (<i>N</i> = 2,918)					
Mixture SDT	5.29	3	.15	8,961	9,002
Unequal variance SDT	9.50	3	.02	8,965	9,007
Equal variance SDT	73.39	4	< .01	9,027	9,063

Note. LR = likelihood ratio goodness-of-fit test; AIC = Akaike's information criterion; BIC = Bayesian information criterion; freq = frequency; SDT = signal detection theory; mixture SDT = Equation 2 in text; unequal variance SDT = Equation 3 in text.

the unequal variance model (Equation 3), and the mixture SDT model (Equation 2). The LR goodness-of-fit statistics show that the equal variance model provides poor fits, whereas the fits for the mixture model and unequal variance model are acceptable. The information criteria for the mixture model are equal to or smaller than those for the unequal variance model in every case, although the difference again tends to be small. Thus, the unequal variance and mixture SDT models again fit about equally well. Note that, for the mixture model, a test of $d_A = 0$ is not rejected in any case; the LR tests give for high frequency 5 s, $LR(1, N = 2,917) = 0.11$, $p = .74$; for low frequency 5 s, $LR(1, N = 2,918) = 0.27$, $p = .60$; for high frequency 2 s, $LR(1, N = 2,816) = 0.21$, $p = .65$; for low frequency 2 s, $LR(1, N = 2,823) = 1.45$, $p = .23$. This again shows that the simple one-parameter extension of Equation 2 appears to be adequate.

Figure 5 shows the data and fitted mixture normal ROC curves on inverse normal coordinates. The curves have slopes less than unity, and are slightly curved but near linear for the range of hits and false alarms obtained in the study. The figure shows that low-frequency words are detected better than high-frequency words, as is usually found, and increasing presentation time increased detection.

The mixture and unequal variance models were also fit to the individual data of the 11 participants in Experiment 5 of Ratcliff et al. (1994), who each participated in four conditions. With respect to fit, the results for 9 of the 11 participants were similar to those shown in Table 3, in that the unequal variance model and the simple mixture model (Equation 2) both provided adequate fits. For example, Table 4 shows results for Participants 2 and 3, who

are representative of most of the others. The table shows that the LR statistics are not significant for most conditions and the values of the AIC are similar across the mixture and unequal variance models (the BIC gave similar results and is omitted for clarity). The LR statistics for 2 other participants (6 and 10), however, were large and significant for all of the conditions, and inspection of the normal ROC curves presented by Ratcliff et al. (1994; Figures 8 and 9) showed that the ROC curves for these participants were markedly nonlinear. In particular, the ROC curves for Participant 6 were curved upward (concave) in all four conditions, whereas those for Participant 10 were curved downward (convex). It is of interest to compare the mixture and unequal variance models for these cases. The lower half of Table 4 shows, for Participants 6 and 10 in all four conditions, the results for fits of the mixture models (Equations 1 and 2) and the unequal variance model; Figure 6 shows some of the data and fitted ROC curves (the results for the other conditions overlapped considerably with those shown and are omitted for clarity).

With respect to Participant 6, Table 4 shows that Equation 1 provides a better fit than either the unequal variance model or Equation 2 for at least three of the four conditions. This reflects the fact that the general mixture model of Equation 1 (with $d_A \neq 0$) can account for upwardly curved normal ROC curves, as shown by the upper left panel of Figure 6, whereas the unequal variance model cannot. This result suggests that the more complex model of Equation 1 might be needed if upwardly curved ROC curves are consistently found for a particular application (an example with a theoretical interpretation is given in DeCarlo, 2000); of course, in this case, additional research is needed to determine whether or not the result is anomalous. With respect to Participant 10, Table 4 shows that, although Equation 1 provides a better fit than the unequal variance model, the LR statistics are still large, mainly because Equation 1 cannot deal with the large downward curvature of the ROC curves, as shown by the upper right panel of Figure 6. However, it is interesting to note that the simple model of Equation 2 with extreme value distributions in place of normal distributions provides good fits of Participant 10's data in all four conditions; the last row of Table 4 presents the results, and the

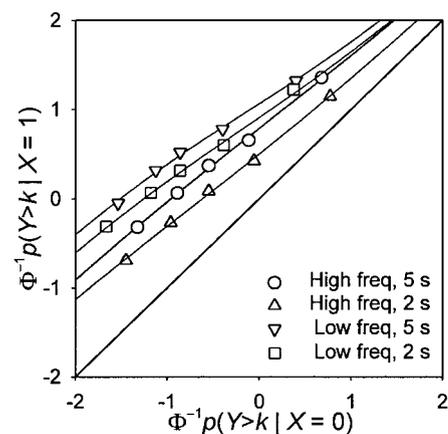


Figure 5. Data and fitted mixture receiver operating characteristic curves on inverse normal coordinates for Experiment 4 of Ratcliff, McKoon, and Tindall (1994). The high- and low-frequency words and their presentation times are indicated in the figure legend. freq = frequency.

Table 4
*Goodness-of-Fit Statistics and Information Criteria for Experiment 5 of Ratcliff, McKoon, and Tindall (1994):
 Participants 2, 3, 6, and 10, Four Conditions*

Model	High freq 1.5 s			Low freq 1.5 s			High freq 5 s			Low freq 5 s		
	LR	p	AIC	LR	p	AIC	LR	p	AIC	LR	p	AIC
Participant 2												
Uneq	3.23	.36	3,003	8.06	.04	2,246	3.24	.36	3,194	6.78	.08	2,408
Mix2	2.46	.48	3,002	5.41	.14	2,243	3.57	.31	3,195	4.25	.24	2,406
Participant 3												
Uneq	15.06	< .01	3,701	5.76	.12	3,341	18.08	< .01	3,593	6.65	.08	2,930
Mix2	4.38	.22	3,690	8.02	.05	3,343	6.54	.09	3,581	18.14	< .01	2,942
Participant 6												
Uneq	20.71	< .01	2,831	7.73	.05	2,985	23.53	< .01	2,805	14.51	< .01	2,839
Mix2	13.47	< .01	2,823	13.88	< .01	2,991	15.41	< .01	2,797	29.22	< .01	2,854
Mix1 ^a	0.99	.61	2,813	3.14	.21	2,983	13.86	< .01	2,797	3.64	.16	2,830
Participant 10												
Uneq	23.76	< .01	2,790	46.07	< .01	2,441	32.71	< .01	2,608	57.69	< .01	2,259
Mix2	15.06	< .01	2,781	31.36	< .01	2,426	22.19	< .01	2,597	25.27	< .01	2,252
Mix1 ^a	9.03	< .01	2,777	27.89	< .01	2,425	20.19	< .01	2,597	21.51	< .01	2,251
Mix2a	2.02	.57	2,768	3.63	.30	2,398	0.93	.82	2,576	1.62	.65	2,229

Note. In all cases except where indicated, *dfs* = 3. freq = frequency; LR = likelihood ratio goodness-of-fit statistic; AIC = Akaike’s information criterion; Uneq = unequal variance SDT (Equation 3 in text); Mix2 = mixture SDT (Equation 2 in text); Mix1 = mixture SDT (Equation 1 in text); Mix2a = Equation 2 in text with extreme value distributions; SDT = signal detection theory.

^a In this row, all *dfs* = 2.

fitted extreme value ROC curves are shown in the lower center panel of Figure 6 (the model uses a log–log link, which corresponds to underlying extreme value distributions that are skewed to the right). Although this is an interesting result, the use of probability distributions other than the normal in the mixture model remains to be investigated; note that, through the use of different link functions, the unequal variance model can be extended in a similar manner (DeCarlo, 1998; Tosteson & Begg, 1988). The mixture model is somewhat more flexible, however, in that it can include more than two latent classes.

Overall, the analysis of individual data shows that the simple mixture model of Equation 2 and the unequal variance model both provide acceptable fits for most of the participants. Table 4 and Figure 6 show that the general mixture model of Equation 1 can also describe the ROC curves for cases where nonlinear normal ROC curves were obtained. Nonlinearity was only found for a few cases, however, so it remains to be determined whether the deviations are indicative about underlying processes or are anomalous. Nevertheless, it is important to note that results that might seem bizarre (to use one reviewer’s description), from the perspective of unequal variance SDT (i.e., the nonlinear ROC curves found for Observers 6 and 10), might be consistent with a systematic process, from the perspective of mixture SDT.

Table 5 presents, for both Experiments 4 and 5, the parameter estimates for the simple mixture model (Equation 2) and the unequal variance model. For Experiment 4, the estimates were obtained by fitting the models to the pooled data given by Ratcliff et al. (1994), whereas for Experiment 5, the models were fit to the

individual data and the table shows the average parameter estimates (across participants). Table 5 shows that, for fits of the mixture model, lower frequency words were detected better than higher frequency words, and detection tended to be better for the 5-s presentation time. The same results were found for fits of the unequal variance model; as before, note that the estimates of the detection parameter are smaller for the unequal variance model than for the mixture model. The right side of Table 5 shows that, for the unequal variance model, the variance tends to be larger for low-frequency words and for the longer presentation time.

The most important result with respect to mixture SDT concerns the parameter λ . The table shows that, for both Experiments 4 and 5, the estimates of λ tend to be larger for the low-frequency words than for the high-frequency words, and the estimates also tend to be larger for the longer presentation times; the latter finding replicates the results found for Experiments 1 and 2 above. Thus, Experiments 4 and 5 of Ratcliff et al. (1994) can be interpreted as showing that more low-frequency words received attention than high-frequency words and that more words were attended to for longer presentation times.

In summary, mixture SDT offers a new framework within which earlier data can be reexamined. For the examples presented above, mixture SDT described the data as well as or better than the unequal variance model. In addition, a comparison of Equations 1 and 2 showed that the simple one-parameter extension of the equal variance model (i.e., Equation 2, with $d_{A'} = 0$) appears to be adequate in most cases. More importantly, the mixture extension of SDT is shown to offer a new interpretation of the data. Con-

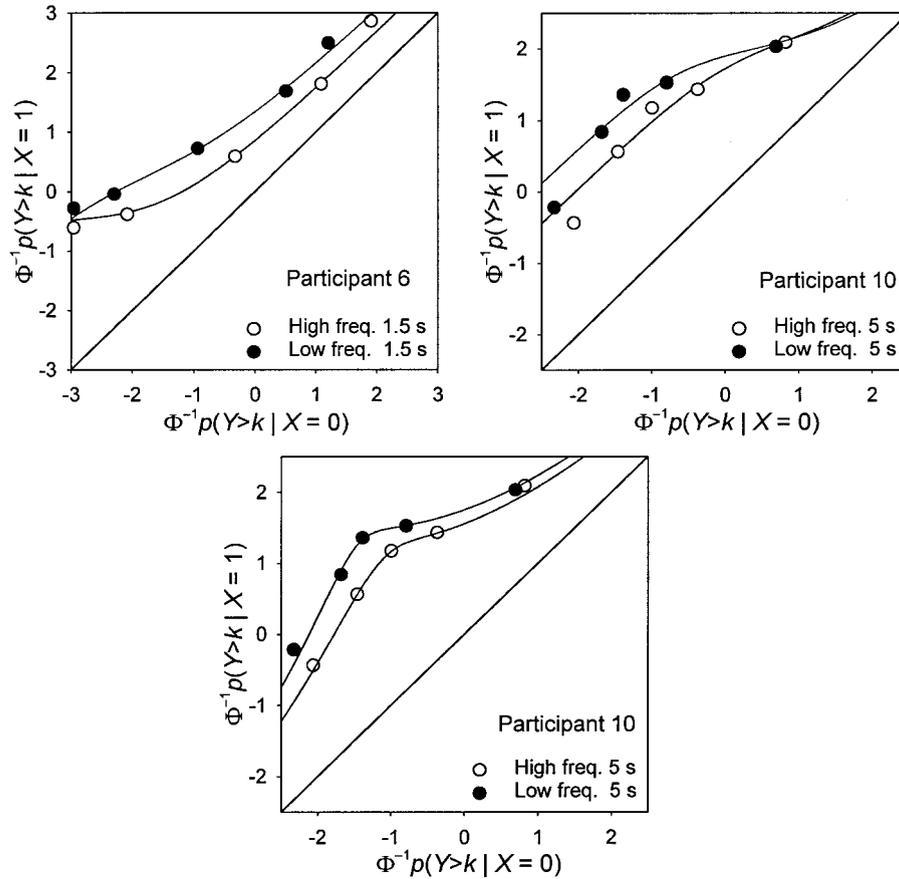


Figure 6. Data and fitted mixture receiver operating characteristic curves for 2 participants in two conditions of Experiment 5 of Ratcliff, McKoon, and Tindall (1994). The upper left and right panels show receiver operating characteristic (ROC) curves for fits of Equation 1; the lower center panel shows ROC curves for a fit of Equation 2 with an extreme value cumulative distribution function in lieu of the normal. freq. = frequency.

sider, for example, the effect of word frequency. Ratcliff et al. (1994) interpreted their data in terms of the unequal variance SDT model and noted that “The larger standard deviation for the low-frequency distribution is plausibly explained by assuming that the familiarity of low-frequency words is more variable for a given subject than the familiarity of high-frequency words” (p. 772). There really is no intuitive reason for why the variance should be larger for low-frequency words, however, as evidenced by the fact that McClelland and Chappell (1998) argued for exactly the opposite view: “We examine the possibility that representations of low-frequency words are less variable than representations of high-frequency words. Support for this possibility comes from the fact that there are far more definitions of high-frequency words than low-frequency words” (p. 743). An interesting aspect of mixture SDT, on the other hand, is that it suggests a different interpretation: Low-frequency words are attended to more often than high-frequency words. Similarly, there is no obvious reason why the variability of a representation should increase as presentation time increases, as found above for fits of the unequal variance model. The interpretation in terms of mixture SDT is again quite simple: More words are attended to with longer presentation times.

Implications for theory. Normal ROC curves for word recognition with slopes less than unity have been interpreted, since Egan (1958), as indicating larger variance for the old-item distribution relative to the new-item distribution. It is important to recognize, however, that the view that nonunit slopes indicate something about the relative variances is an interpretation in terms of the unequal variance extension of SDT and is not a result. From the perspective of mixture SDT, nonunit slopes do not provide information about the relative variances but rather about the mixture process. Note that, from the perspective of unequal variance SDT, what appears to be a systematic increase in the signal variance as detection increases arises quite naturally from a mixture process, because mixing distributions (i.e., signal and noise) that are farther apart results in a distribution that appears to have larger variance.

Mixture SDT provides an alternative account of nonunit slope ROC curves; it also has implications about the way new versions of global memory models are formulated (cf. Ratcliff et al., 1994, p. 779). In particular, instead of adding or modifying assumptions about the variance, mixture SDT suggests that another approach is to consider processes that might lead to latent classes of items, such as attention (and most current theories already include assumptions about attention).

Table 5
Parameter Estimates for Experiments 4 and 5 of Ratcliff, McKoon, and Tindall (1994)

Condition	Mixture SDT		Unequal variance SDT		
	d_A	λ	d_N	$\log \sigma_s$	σ_n/σ_s
Experiment 4					
High freq 2 s	1.32 (0.15)	0.47 (0.05)	0.61 (0.05)	0.20 (0.04)	0.82
Low freq 2 s	1.69 (0.10)	0.70 (0.04)	1.25 (0.06)	0.29 (0.04)	0.75
High freq 5 s	1.32 (0.11)	0.70 (0.05)	0.95 (0.05)	0.19 (0.04)	0.83
Low freq 5 s	1.86 (0.09)	0.76 (0.03)	1.51 (0.06)	0.35 (0.04)	0.70
Experiment 5					
High freq 1.5 s	1.85 (0.11)	0.67 (0.04)	1.28 (0.10)	0.33 (0.04)	0.72
Low freq 1.5 s	2.28 (0.06)	0.83 (0.02)	2.09 (0.07)	0.42 (0.05)	0.66
High freq 5 s	2.11 (0.14)	0.77 (0.04)	1.76 (0.13)	0.42 (0.07)	0.66
Low freq 5 s	2.67 (0.12)	0.90 (0.02)	2.95 (0.18)	0.56 (0.06)	0.57

Note. Standard errors are in parentheses. Experiment 4: pure conditions, pooled data. Experiment 5: pure conditions, mean and standard error of the parameter estimates obtained from individual fits for 11 participants in four conditions (except $n = 10$ for the low freq 5 s condition, mixture SDT, because of convergence problems for the analysis of 1 participant). SDT = signal detection theory; mixture SDT = Equation 2 in text; unequal variance SDT = Equation 3 in text; freq = frequency.

Although the focus here has been on recognition memory, mixture SDT models can be readily motivated for other memory phenomena as well, such as the effects of repetition, source discrimination (models for repetition and source discrimination were presented in DeCarlo, 2000), the mirror effect, and the process dissociation procedure (which readily suggests a mixture process). Repetition is an interesting example in that it suggests a situation where three or more latent classes might be needed. Consider what might happen, for example, when items are presented twice. If it is assumed that an item attended to twice is shifted further than an item attended to once,² then it follows from the mixture point of view that there will be a total of three latent classes of items on signal trials: a proportion of items attended to twice, a proportion attended to once, and a proportion not attended to on either presentation. Thus, there will be at least two distributions with nonzero locations corresponding to once- and twice-attended items (with the nonattended distribution assumed to have the same location as the new item distribution, as in Equation 2). Some evidence for this was obtained for an analysis of data of Egan (1958) in that a fit of the mixture model suggested that repetition of a study list resulted in two distributions with nonzero locations, whereas one repetition gave one nonzero distribution (DeCarlo, 2000); the two distributions can be interpreted as items attended to once and twice. In contrast, the unequal variance model suggested that repeating the list did not affect the variance of the signal distribution, which raises questions as to why other variables that increase strength also increase the variance but not repetition. This led Glanzer et al. (1999) to note that "It may be that repetition differs intrinsically from other variables" (p. 508). Mixture SDT suggests an interesting possibility as to how and why it might differ: Repetition gives rise to additional latent classes of items. This remains to be investigated.

Conclusions

SDT can be extended in a psychologically meaningful way by allowing for mixtures of the underlying distributions. The mixtures

are viewed as arising from the action of a second process, with the result that signal trials consist of latent classes of trials, with the mixing parameter indicating the proportion of each type of trial. It is shown that this simple idea can be applied to a variety of situations and offers new insights into the data. Examples in recognition memory research are given where mixture SDT offers a new account of the effects of various variables, and the results suggest new directions for future research.

Although recognition memory research is considered here, the approach is general and can be applied to any area of research where SDT has been applied. Consider, for example, Swets's (1986) observation that normal ROC curves for odor recognition have slopes close to unity. An interpretation of this result in terms of mixture SDT is that all of the odors received attention during the study period, perhaps reflecting an intrinsic difference between odors (they command attention?) and visual stimuli such as words. In this way, the mixture perspective suggests new interpretations of old data.

Although only mixing on signal trials was considered here, situations where there might be mixtures on noise trials or on both signal and noise trials can also be motivated (an example was discussed in DeCarlo, 2000). For example, Swets (1986) noted that slopes greater than unity were found for detection of abnormal cells (presented on slides). From the mixture perspective, a slope greater than unity could result from a mixture of noise distributions; this suggests that there might have been two latent classes of slides for nondiseased cases—perhaps a proportion of the nondiseased slides had signal-like aspects (spots or discolorations), which suggests reexamining the nondiseased slides to see if this might be the case. Mixture SDT provides a framework for investigating this and other possibilities; it greatly enriches the theoretical scope of

² Another approach is to assume that the effect of attention is all or none, so that attending twice does not further shift the distribution; the resulting model is then simply Equation 2.

SDT, opening new possibilities for applications in psychology and other areas.

References

- Agresti, A. (1990). *Categorical data analysis*. New York: Wiley.
- Burnham, K. P., & Anderson, D. R. (1998). *Model selection and inference: A practical information-theoretic approach*. New York: Springer-Verlag.
- Dayton, C. M. (1998). *Latent class scaling analysis*. Thousand Oaks, CA: Sage.
- DeCarlo, L. T. (1998). Signal detection theory and generalized linear models. *Psychological Methods*, 3, 186–205.
- DeCarlo, L. T. (2000, August). *An extension of signal detection theory via finite mixtures of generalized linear models*. Paper presented at the 2000 meeting of the Society for Mathematical Psychology, Kingston, Ontario, Canada. Available at <http://www.columbia.edu/~ld208>.
- DeCarlo, L. T. (in press). A latent class extension of signal detection theory, with applications. *Multivariate Behavioral Research*.
- Egan, J. P. (1958). *Recognition memory and the operating characteristic*. (Tech. Note No. AFCRC-TN-58-51, AO-152650). Indiana University Bloomington, Hearing and Communication Laboratory.
- Gescheider, G. A. (1997). *Psychophysics: The fundamentals*. Hillsdale, NJ: Erlbaum.
- Glanzer, M., Adams, J. K., Iverson, G. J., & Kim, K. (1993). The regularities of recognition memory. *Psychological Review*, 100, 546–567.
- Glanzer, M., Kim, K., Hilford, A. M., & 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.
- Greene, W. H. (1998). *LIMDEP Version 7.0 user's manual* (Rev. ed.). New York: Econometric Software.
- Macmillan, N. A., & Creelman, C. D. (1991). *Detection theory: A user's guide*. New York: Cambridge University Press.
- McClelland, J. L., & Chappell, M. (1998). Familiarity breeds differentiation: A subjective-likelihood approach to the effects of experience in recognition memory. *Psychological Review*, 105, 724–760.
- McLachlan, G., & Peel, D. (2000). *Finite mixture models*. New York: Wiley.
- Muthén, L. K., & Muthén, B. O. (1998). *Mplus user's guide*. Los Angeles: Authors.
- Ratcliff, R., McKoon, G., & Tindall, M. (1994). Empirical generality of data from recognition memory receiver-operating characteristic functions and implications for the global memory models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20, 763–785.
- Ratcliff, R., Sheu, C. F., & Gronlund, S. D. (1992). Testing global memory models using ROC curves. *Psychological Review*, 99, 518–535.
- Ratcliff, R., Van Zandt, T., & McKoon, G. (1995). Process dissociation, single-process theories, and recognition memory. *Journal of Experimental Psychology: General*, 124, 352–374.
- Swets, J. A. (1986). Form of empirical ROCs in discrimination and diagnostic tasks: Implications for theory and measurement of performance. *Psychological Bulletin*, 99, 181–198.
- Swets, J. A. (1996). *Signal detection theory and ROC analysis in psychology and diagnostics: Collected papers*. Hillsdale, NJ: Erlbaum.
- Tosteson, A. N. A., & Begg, C. B. (1988). A general regression methodology for ROC curve estimation. *Medical Decision Making*, 8, 204–215.
- Vermunt, J. K. (1997). LEM: A general program for the analysis of categorical data [Computer software and manual]. Retrieved from <http://www.uvt.nl/faculteiten/fsw/organisatie/departementen/mto/software2.html>
- Yonelinas, A. P. (1994). Receiver-operating characteristics in recognition memory: Evidence for a dual-process model. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20, 1341–1354.
- Yonelinas, A. P. (1999). The contribution of recollection and familiarity to recognition and source-memory judgments: A formal dual-process model and an analysis of receiver operating characteristics. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25, 1415–1434.

Appendix

Some Notes on Fitting and Testing Mixture Models

Detection

Equations 1 and 2 can be fit using software for mixture models, such as LEM (Vermunt, 1997), which is available at <http://www.uvt.nl/faculteiten/fsw/organisatie/departementen/mto/software2.html>. LEM allows one to fit a mixture multinomial distribution with generalized linear models in each component of the mixture. For example, Equation 2 can be fit by writing it as a latent class model with an observed covariate X ,

$$p(Y \leq k|X) = \sum_w p(W) p(Y \leq k|XW),$$

where W is a latent dummy (zero/one) variable, $p(W)$ is the mixing proportion (λ in Equation 2), and the last term can be written as a cumulative probit model:

$$p(Y \leq k|XW) = \Phi(c_k - d_A XW).$$

The model is specified in LEM by the terms $p(W)$ and $p(A|XW)$, with a cumulative probit model specified for $p(A|XW)$; a sample program is available from my Web site (<http://www.columbia.edu/~ld208>). Note that it is important to run the program repeatedly with different starting values, because local maxima are sometimes encountered; see Dayton (1998) for

a general discussion of various aspects of latent class analysis and DeCarlo (in press) for discussion of a latent class extension of signal detection theory (SDT) and additional sample programs.

To fit Equation 1, X is included as a predictor in addition to XW , and the model is

$$p(Y \leq k|XW) = \Phi(c_k - d_A X - d_{AA} XW),$$

with $W = 0$ indicating the nonattended trials, $W = 1$ the attended trials, and d_{AA} indicating the distance of the A distribution from the A' distribution; d_A is obtained as $d_{AA} + d_{A'}$. Note that which class serves as the reference is arbitrary, but it can be determined from the output. The class corresponding to lack of attention in Equation 2, for example, is obvious from the conditional probabilities provided in the output, because they are equal across signal and noise. For Equation 1, the signs of the two coefficients (and inspection of the conditional probabilities) allow one to determine which class corresponds to A' and which to A (responses are coded so that higher numbers indicate higher confidence)—both coefficients negative indicates that the coefficient of XW gives d_{AA} and the coefficient of X gives $d_{A'}$ (i.e., $W = 1$ indicates attention, as used above), whereas positive and negative coefficients indicate that the coefficient of XW gives d_{AA} and

the coefficient of X gives d_A (i.e., $W = 0$ indicates attention). The solutions are equivalent, in that they simply redefine the reference distribution (as noise or signal), and the parameter estimates are the same. Note that the standard errors might have to be computed; that is, if one wants the standard error of d_A from output where the standard errors of $d_{AA'}$ and $d_{A'}$ are given, then it can be obtained as (the square root of) $\text{Var}(d_A) = \text{Var}(d_{A'} + d_{AA'}) = \text{Var}(d_{A'}) + \text{Var}(d_{AA'}) + 2 \text{Cov}(d_{A'}, d_{AA'})$. Including the command "wse filename" in LEM will write out the estimates of the variances and covariances of the parameter estimates to a file specified by filename.

The mixture SDT model (with logistic distributions) can also be fit using software for second-generation structural equation modeling, such as Mplus (Muthén & Muthén, 1998). A sample program for mixture SDT is available at my Web site.

The Unequal Variance SDT Model

Equation 3 is more generally known in econometrics and biostatistics as an ordered probit model with multiplicative heteroscedasticity; it was fit here using the software LIMDEP (Greene, 1998), which gives maximum-likelihood estimates and also allows one to fit more general models with heteroscedasticity by using different link functions. The model (i.e., Equation 3) is implemented in LIMDEP as

$$p(Y \leq k|X) = \Phi[(c_k - d_N X) / \exp(\tau X)],$$

where \exp is the exponential function; the use of the exponential function ensures that the estimate of the variance is positive. Note that the above is identical to Equation 3 with $\tau = \ln(\sigma_s)$; thus, as shown in the tables, the fitted model gives estimates of (natural) $\log(\sigma_s)$ and its standard error.

A Note on Log-Likelihoods

The maximized log-likelihoods are used to compute the information criteria; here it is noted that the log-likelihoods reported by LEM and LIMDEP differ by a constant. Let A indicate the response, X the signal or noise, and n the frequency of pattern AX . The log-likelihood reported by LEM can then be written as

$$\log L = \sum n \log p(AX),$$

where $p(AX)$ is the joint probability of A and X , and the summation is over the different patterns of AX . In a similar manner, the log-likelihood given by LIMDEP is

$$\log L = \sum n \log p(A|X),$$

where $p(A|X)$ is the conditional probability of A given X . Note that

$$\begin{aligned} \sum n \log p(AX) &= \sum n \log [p(A|X)p(X)] \\ &= \sum n \log p(A|X) + \sum n \log p(X), \end{aligned}$$

which shows that subtracting $\sum n \log p(X)$ from the log-likelihood reported by LEM gives the log-likelihood reported by LIMDEP, and this was done to compute the information criteria reported in the tables. Note that, for a fit of the equal variance model, the log-likelihood obtained in this manner by LEM was identical in every case to three decimal places to that given by LIMDEP.

Received August 1, 2000

Revision received November 5, 2001

Accepted November 5, 2001 ■