Using the PLUM procedure of SPSS to fit unequal variance and generalized signal detection models

LAWRENCE T. DECARLO
Teachers College, Columbia University, New York, New York

The recent addition of a procedure in SPSS for the analysis of ordinal regression models offers a simple means for researchers to fit the unequal variance normal signal detection model and other extended signal detection models. The present article shows how to implement the analysis and how to interpret the SPSS output. Examples of fitting the unequal variance normal model and other generalized signal detection models are given. The approach offers a convenient means for applying signal detection theory to a variety of research.

Signal detection theory (SDT) has been widely used in psychology and other fields (Gescheider, 1997; Macmillan & Creelman, 1991; Swets, 1996). A signal detection model is basically a model of processes involved when observers make a decision as to whether or not an event occurred or which of two or more events occurred, to name just a few possibilities. An attractive aspect of the theory, from a psychological perspective, is that it separates arbitrary decision factors (namely, the placement of response criteria on an underlying dimension) from perceptual or memorial factors, for example (namely, the ability of an observer to detect or remember). It has long been recognized that observers differ in arbitrary ways with respect to their use of such response categories as *very sure* versus *sure*, for example, and so the separation in SDT of response criteria from detection is important.

The equal variance signal detection model can easily be fit using standard statistical software; sample SPSS and SAS programs are given in DeCarlo (1998). Up to this point, however, extensions of the basic SDT model have been more difficult to fit. For example, the unequal variance extension of the SDT model is often fit using specialized software (e.g., ROCKIT; Metz, 1998). The normal SDT model with unequal variances is equivalent to a model more generally known in statistics as a probit model with heteroscedastic error, and so the model can also be fit using statistical software such as STATA (StataCorp, 2001) or econometric software such as LIMDEP (Greene, 1998; for an application in SDT, see DeCarlo, 2002). Researchers in psychology and education, however, are generally not familiar with these specialized and somewhat more advanced software packages. The unequal variance SDT

model can also be fit using software for nonlinear models, as was noted by DeCarlo (1998), and Sheu and Heathcote (2001) have recently provided code to fit the unequal variance normal SDT model using PROC NONLIN of SAS. This still puts the model beyond the reach, however, of many researchers who are not familiar with the SAS language.

The recent addition of a procedure in SPSS for the analysis of ordinal regression models—namely, the PLUM (polytomous universal model) procedure—enables researchers to fit a variety of signal detection models, including the unequal variance model, by simply pointing and clicking. This is important because applied researchers in psychology and education, who could greatly benefit from a signal detection approach, can now perform the analysis by using software they are already familiar with. Of course, one still has to understand how the models are parameterized in SPSS. The present article shows how to implement the analysis and how to interpret the SPSS output. The basic equal variance normal SDT model is presented first, followed by the unequal variance extension. The steps needed to fit the model in SPSS are given, and an application to a classic dataset is shown. One can also fit generalized signal detection models (DeCarlo, 1998) with the PLUM procedure by using different link functions, as will be shown below and illustrated with an example. Finally, a more complex example that involves several covariates and interaction terms is presented.

SIGNAL DETECTION THEORY AND ORDINAL REGRESSION MODELS

The Unequal Variance Normal SDT Model

The equal variance normal theory SDT model for rating or binary responses can be written as

$$p\left(Y \leq k \mid X\right) = \Phi\left(\frac{c_k - dX}{\sigma}\right), \tag{1}$$

for $k = 1$ to $K-1$, where K is the number of response categories, Y is a response variable (e.g., a confidence rating)

Correspondence should be addressed to L. T. DeCarlo, Department of Human Development, Box 118, Teachers College, Columbia University, 525 West 120th Street, New York, NY 10027 (e-mail: decarlo@ exchange.tc.columbia.edu). Additional programs for signal detection analysis are available at the author's Web site, http://www.columbia. edu/~ld208.

that takes on values $k = 1$ to K, X is a dummy coded variable (e.g., noise = 0, signal = 1), $p(Y \leq k|X)$ is the cumulative probability of a response of k or less given X, Φ is the cumulative normal distribution function, c_k are response criteria (distances from the noise distribution) with the property $c_1 < c_2 < \ldots < c_{K-1}$, d is the distance of the signal distribution from the noise distribution (for normal distributions, this is simply d', but the more general notation used in DeCarlo, 1998, is used here), and σ is the standard deviation of the underlying distributions, which can be set to unity without loss of generality.

The basic SDT model shown in Equation 1 is more generally known as a probit model; the relation of the model to ordinal regression models used in statistics has previously been discussed by DeCarlo (1998). The present article focuses on other extensions of the basic model. For example, the most widely used extension allows the variances of the underlying distributions to differ across signal and noise (Green & Swets, 1966), which gives the unequal variance normal SDT model,

$$p\left(Y \leq k \middle| X\right) = \Phi\left(\frac{c_k - d_n X}{\sigma_s^X}\right), \qquad (2)$$

where d_n is the detection parameter scaled with respect to the standard deviation of the noise distribution (σ_n, which is set to unity), and σ_s is the standard deviation of the signal distribution. Equation 2 generalizes the equal variance normal SDT model with the addition of the parameter $\sigma_s > 0$; note that $1/\sigma_s$ gives the ratio of the standard deviations of the noise and signal distributions, which corresponds to the slope of the receiver operating characteristic (ROC) curve on inverse normal coordinates. Given that $\sigma_s > 0$, it is convenient to rewrite the denominator of the above using the exponential function,

$$p\left(Y \leq k \middle| X\right) = \Phi\left(\frac{c_k - d_n X}{\exp(aX)}\right), \qquad (3)$$

where exp is the exponential function and $a = \ln \sigma_s$ (where ln is the natural logarithm). Applying the inverse normal transform Φ^{-1} (i.e., a probit transform) to both sides of the above gives

$$\Phi^{-1}\left[p\left(Y \leq k \middle| X\right)\right] = \frac{c_k - d_n X}{\exp(aX)}, \qquad (4)$$

which expresses the model in the form used in the PLUM procedure of SPSS, which is a part of the advanced models module. The model is more generally known in statistics and econometrics as an ordered probit model with (multiplicative) heteroscedastic error (see, e.g., Greene, 2000); McCullagh and Nelder (1989) discussed a general version of Equation 4 based on the logistic distribution (i.e., a logit model with heteroscedastic error); Tosteson and Begg (1988) discussed a general version of Equation 4 based on generalized linear models (see below). Note that Equation 4 can easily be extended to situations involving more than one signal, say M signals, by replacing

$d_n X$ by $\sum_i d_{ni} X_i$ for $i = 1$ to M and $\exp(aX)$ by $\exp(\sum_i a_i X_i)$; see Example 3 below. Also, covariates included in the denominator do not have to be the same as those included in the numerator (as in Equation 4); that is, for the unequal variance SDT model, the same dummy coded signal indicators are included in the numerator and the denominator to allow the locations and variances of the underlying distributions to differ; however different covariates can also be included in either the numerator or the denominator, as will be shown below.

With respect to the SPSS output, d_n is referred to as a location parameter and a as a scale parameter; the above shows that, for the normal model, exponentiating the estimated scale parameter and taking the inverse gives an estimate of σ_n/σ_s, which is the slope of the inverse normal ROC curve.

Example 1: Unequal Variance Normal SDT

To illustrate use of the PLUM procedure, the data for Observer 1 from a light detection study of Swets, Tanner, and Birdsall (1961) are analyzed; the data have been widely analyzed and previously used to illustrate the application of generalized SDT models (DeCarlo, 1998). The data, as given by Green and Swets (1966, p. 102), consist of a 2 × 6 table, with the two rows corresponding to signal or noise presentations and the six columns to the 1–6 rating response. Table 1 shows the data and how they are set up in the SPSS file. There are three columns of variables: a dummy coded variable X that indicates whether noise (0) or signal (1) was presented, a response variable Y that takes on values from 1 to 6 for the six confidence rating categories, and a variable that indicates the frequency of each response to each stimulus. Because the data are in tabular form, a weighted analysis must be performed (to recognize the response frequencies). This can be done in SPSS by choosing **data** followed by **weight cases** and indicating that the frequency variable is a weight. In most cases, however, researchers will have data in the form of individual records, in which case one can proceed directly to the analysis without weighting cases.

Next, choose **analyze**, **regression**, and **ordinal**. Choose Y as the dependent variable and enter the signal indicator

Table 1
Data as Organized for SPSS: Observer 1
From Swets, Tanner, and Birdsall (1961)

X	Y	Frequency
0	1	174
0	2	172
0	3	104
0	4	92
0	5	41
0	6	8
1	1	46
1	2	57
1	3	66
1	4	101
1	5	154
1	6	173

Table 2
Goodness-of-Fit Statistics and Information Criteria for
Observer 1 From Swets, Tanner, and Birdsall (1961), $N = 1{,}188$

Model	LR	df	p	X^2	df	p	BIC	AIC
Equal variance normal	32.972	4	<.01	33.728	4	<.01	3,915.94	3,885.46
Unequal variance normal	1.482	3	.69	1.497	3	.68	3,891.53	3,855.97
Equal variance extreme value	5.125	4	.28	5.574	4	.23	3,888.09	3,857.61

Notes—LR, likelihood ratio goodness-of-fit statistic; X^2, Pearson goodness-of-fit statistic; AIC, Akaike's information criterion; BIC, Bayesian information criterion.

X as a covariate. Note that if X is entered as a factor, effect coding (-1 and 1) will be used, in which case the location parameter is one half of the detection parameter and the criteria are redefined as the distances from the intersection point of the two underlying normal distributions (see De-Carlo, 1998; note that the denominator of Equation 4 is also affected); the recommendation here is to simply code the signal indicator as zero/one and enter it as a covariate. Under **options**, choose the probit link in order to fit an SDT model with normal distributions (the other links will be discussed below). Under **output**, it is useful to change the **print log-likelihood** option to **excluding multinomial constant**, because the printed log-likelihood is then the same as that reported in other articles (e.g., DeCarlo, 1998, 2002). Up to this point, the model being fit is an equal variance normal SDT model. As a first step, the reader should fit this model and compare the results with those given in the first row of Table 2 (discussed below).

To fit the unequal variance normal SDT model, choose **scale** and enter the dummy coded signal indicator X as a main effect. This gives the denominator shown in Equation 4 above. The resulting syntax is

```
PLUM
y WITH x
/CRITERIA = CIN(95) DELTA(0) LCONVERGE(0)
MXITER(100) MXSTEP(5)
PCONVERGE(1.0E-6) SINGULAR(1.0E-8)
/LINK = PROBIT
/PRINT = FIT KERNEL PARAMETER SUMMARY
/SCALE =x
```

Note that omitting the last line from the above gives the equal variance normal SDT model.

Table 2 shows likelihood ratio (LR) and Pearson goodness-of-fit tests reported in the SPSS output in a table labeled "Goodness-of-Fit" (note that the LR test is referred to as the deviance in the SPSS output), along with the degrees of freedom and probability values. Both fit statistics are asymptotically distributed as a chi-square, but they can differ in some situations, such as when there are small counts in many cells of the table (i.e., the table is sparse),

in which case they might not follow the chi-square distribution and should not be trusted (see McCullagh & Nelder, 1989). The goodness-of-fit statistics test the null hypothesis that the model fits the data. The first row of the table shows that, in terms of fit, the equal variance normal SDT model is rejected by both tests. The second row of the table shows that the unequal variance normal SDT model, on the other hand, is not rejected. The third row of the table and the information criteria (AIC and BIC) will be discussed in the next section. Overall, Table 2 shows that the unequal variance normal SDT model adequately describes the data.

Table 3 shows estimates of the parameters and standard errors (SEs) for a fit of the unequal variance normal SDT model. The estimates of c_k represent the distances of the response criteria from the mean of the noise distribution (which serves as the zero point). Table 3 shows that the estimate of the detection parameter d_n is 1.519 with a standard error of 0.096; note that Dorfman and Alf (1969) reported an estimate of Δm (which is the same as d_n) for this observer of 1.51. The SPSS output also provides 95% confidence intervals for the parameters, which are computed by multiplying the estimates of the SEs by 1.96 and adding and subtracting the result to the point estimates. The confidence intervals are useful for comparing detection across different groups or conditions.

The estimate of the scale parameter a (ln σ_s) and its SE are also shown in Table 3. As was noted above, the estimate of the ratio of the noise to signal standard deviations is obtained by exponentiating the scale parameter a and taking the inverse (or simply exponentiating $-a$), which for Observer 1 gives an estimate of $\exp(-0.348) = 0.71$, which is identical to the value reported by Dorfman and Alf (1969; the ratio is given by their parameter b). Note that the output gives an estimate of the SE of the scale parameter $a = \ln \sigma_s$, and not an estimate of the SE of $1/\sigma_s$ (i.e., the SE for the slope of the ROC curve). A Taylor series expansion can be used to obtain an approximate estimate of the SE of $1/\sigma_s$; however, it is not really needed, in that the SE and confidence intervals for a, which are given

Table 3
Parameter Estimates (With Standard Errors) for the Unequal Variance Normal SDT Model (Equation 3)
for Observer 1 From Swets, Tanner, and Birdsall (1961)

d_n		ln σ_s		c_1		c_2		c_3		c_4		c_5	
Est.	SE	Est.	SE	Est.	SE	Est.	SE	Est.	SE	Est.	SE	Est.	SE
1.519	0.096	.348	0.063	-0.533	0.054	0.204	0.050	0.710	0.053	1.366	0.067	2.294	0.113

in the SPSS output, essentially provide the same information (with respect to comparing the variances across groups or conditions, for example).

The SPSS output also provides a table labeled "Model Fitting Information." The chi-square statistic shown in this table provides a test of an intercept-only model (i.e., a model with response criteria and no other predictors) against a final model (i.e., the model of interest). For example, for a fit of an equal variance SDT model, the chi-square statistic provides a test of the null hypothesis that the detection parameter is equal to zero. On the other hand, for a fit of an unequal variance SDT model, the null hypothesis being tested is that both the detection parameter and the scale parameter are zero (i.e., zero detection and unit variance). The most useful aspect of the table is that it shows estimates of the minus two log likelihood ($-2LL$) for the final and intercept-only models; the $-2LL$ can be used to compute information criteria and in tests of nested models, as will be shown below.

The output also provides a table labeled "Pseudo R-Square" and reports three measures, labeled as Cox and Snell (1989), Nagelkerke (1991), and McFadden (1974). The measures have not been used in SDT research; Agresti (1990) noted that none of the measures appears to be as useful as R^2 in ordinary regression. For a discussion of pseudo-R^2 measures in the context of categorical data analysis, see Agresti, Greene (2000), Maddala (1983), or Menard (2000); note that the different measures arise because there are different ways to define the variance for categorical data.

Generalized SDT Models

The above section has shown that the unequal variance normal SDT model can easily be fit using SPSS. Here, it is noted that more general SDT models with nonnormal underlying distributions can also be fit; these have been referred to as *generalized signal detection models*, in that the generalization is accomplished through the use of generalized linear models (see DeCarlo, 1998, for details and sample SAS programs). In particular, note that the equal variance normal model given by Equation 1 can be generalized by using, on the right side, other cumulative distribution functions (CDF) in place of the normal,

$$p\left(Y \leq k \,|\, X\right) = F\left(\frac{c_k - dX}{\tau}\right),$$

where F is a CDF and τ is a scale parameter (i.e., the detection and criteria parameters are scaled differently for different distributions; see DeCarlo, 1998). Applying the inverse of F to both sides of the above gives the model written as a generalized linear model,

$$g\left[p\left(Y \leq k \,|\, X\right)\right] = \frac{c_k - dX}{\tau}, \tag{5}$$

where $g = F^{-1}$ is known as a *link function* (McCullagh & Nelder, 1989). In short, in the context of SDT, using a different link function gives an SDT model based on different underlying distributions, in that the inverse of the link function corresponds to a CDF.

With respect to the PLUM procedure, five different link functions are offered: probit, logit, complementary log-log, negative log-log, and cauchit links. Let γ_k be the cumulative response probability $p(Y \leq k \,|\, X)$. The probit link, $\Phi^{-1}(\gamma_k)$, was used above for the normal SDT model. The logit link, $\ln[\gamma_k / (1 - \gamma_k)]$, gives an SDT model based on logistic underlying distributions. The complementary log-log link, $\ln[-\ln(1 - \gamma_k)]$, gives a model based on extreme value distributions (the distribution of smallest extremes, which is skewed to the left); for examples of applications of this model, see DeCarlo (1998). The negative log-log link, $-\ln[-\ln(\gamma_k)]$, also gives a model based on extreme value distributions; however, it is a distribution of largest extremes, which is skewed to the right. The cauchit link, $\tan[\pi(\gamma_k - 0.5)]$, gives a model based on the Cauchy distribution, which has heavy tails (and a kurtosis of infinity; see DeCarlo, 1997). Thus, signal detection models based on a variety of underlying distributions can easily be examined.

A somewhat overlooked but useful aspect of the approach via generalized SDT models is that, through the use of different link functions, one can examine the robustness of results (e.g., the estimates of d across different conditions) with respect to the form of the underlying distribution; one can also obtain information about how the underlying distributions might deviate from normality. For example, the logit and cauchit links give SDT models based on distributions with heavier tails than the normal (with the Cauchy having heavier tails than the logistic), whereas the complementary log-log and minus log-log links give SDT models based on distributions that are negatively and positively skewed, respectively. Thus, one can easily consider skewed and heavy-tailed underlying distributions in lieu of the normal distribution via the different link functions. The absolute and relative fits of the models can be assessed by using goodness-of-fit statistics and information criteria, as will be discussed next.

Example 2: An Extreme Value SDT Model

It has previously been shown (DeCarlo, 1998) that a model based on the extreme value distribution (of smallest extremes) also describes the data of Swets et al. (1961) analyzed above. The analysis is the same as the above, except that a complementary log-log link is used in lieu of a probit link, and an equal variance model is fit (i.e., the dummy signal variable is removed from the scale option). Note that the parameter estimates, standard errors, fit statistics, and minus two log likelihood given in the SPSS output in this case are the same as those given in Table 1 of DeCarlo (1998; where SAS was used). The third row of Table 2 shows the results with respect to fit. The goodness-of-fit statistics show that the equal variance extreme value SDT model is not rejected.

Table 2 also shows two information criteria, a Bayesian information criterion (BIC) and Akaike's information criterion (AIC; see Agresti, 1990; Burnham & Anderson, 1998); the information criteria are useful for comparing nonnested models. In the context of SDT, models based on different underlying distributions are not nested, in that

one model cannot be obtained by restricting parameters of the other model, and so a likelihood ratio test cannot be used to compare the models. Information criteria, however, can be used to compare the models, with smaller values indicating a better model. (For some examples of using information criteria to compare the unequal variance SDT model with mixture SDT models, see DeCarlo, 2002.) Here, it is noted that the information criteria are easily computed from the SPSS output. Specifically, the criteria can be computed as follows, using the estimates of the $-2LL$ given in the SPSS output,

$$AIC = -2LL + 2 \times \# \, par \qquad (6)$$

$$BIC = -2LL + \ln(N) \times \# \, par, \qquad (7)$$

where N is the sample size, ln is the natural logarithm, and # par is the number of parameters.

With respect to the data of Swets et al. (1961), the BIC in Table 2 is smallest for the equal variance extreme value SDT model, whereas the AIC is smallest for the unequal variance normal SDT model. Note that the equal variance extreme value model has one less parameter than the unequal variance normal model, and so, as compared with the AIC, the BIC in this case favors a model with fewer parameters (although the differences are small). Dayton (1998) noted that, because for $N \geq 8$ the "penalty" term for the number of parameters is larger for BIC than for AIC (i.e., $\ln(N) \times \#$ par for BIC and $2 \times \#$ par for AIC), the BIC tends to select models with fewer parameters than does the AIC, and the present example appears to be a case in point.

Together, the goodness-of-fit statistics and information criteria in Table 2 suggest that, for this example, one cannot really choose between the models solely on the basis of statistical criteria, in that both models describe the data and there is no clear-cut evidence for one model over the other. What is needed are other types of evidence (such as experimental evidence) for or against the validity of the model; for an example of this approach in the context of comparing the unequal variance normal SDT model with a normal mixture extension of SDT, see DeCarlo (2002).

It is also useful to look at ROC plots of the data along with fitted ROC curves (for examples, see DeCarlo, 1998, 2002). In SPSS, an ROC plot on probability coordinates and a nonparametric measure of detection can be obtained by using the **graphs** option. In my view, it is more informative to examine plots on transformed (i.e., linearizing) coordinates, since it is easier to eye a straight line than a curve. For the normal SDT model, the plots can be constructed by using a probit function to transform the obtained proportions; the function is available in the SPSS base language. The transforms for logistic or extreme value SDT models (or Cauchy models), which are shown in DeCarlo (1998), can also easily be computed.

One can also consider more general SDT models by using different link functions and allowing the scale factor in Equation 5 to differ across signal and noise; this gives a general class of unequal variance SDT models discussed by Tosteson and Begg (1988). With respect to the

data of Swets et al. (1961), if a scale factor is included in the extreme value model, then the estimate of a does not differ significantly from zero, which indicates that the variance of the underlying extreme value distribution is the same across signal and noise. General SDT models with different link functions and unequal variances have generally not been used in psychology and so remain to be investigated. It should also be noted that unequal variance extensions of signal detection models raise some issues, and other generalizations of equal variance SDT models have been proposed (DeCarlo, 2000, 2002).

The next section will present a more complex example that further illustrates how to use the PLUM procedure to perform signal detection analysis. It is shown, for example, that hypotheses about differences across groups of participants can be tested by including interaction terms in the location or scale parts of the model. Of course, one has to pay attention to exactly how the covariates are introduced. The conceptualization via SDT is useful in this regard in that each effect has a specific interpretation in terms of the underlying theory.

Example 3: Inclusion of Additional Covariates

The data are from a pilot study concerned with racial attitudes, memory, and perception of symptom severity (Gushue, 2002). Eighteen participants, who were white graduate students in a counseling and clinical psychology program, read a one-page clinical description of a client named Rob; 10 of the participants were told that the client was black, and 8 participants were told that the client was white. After reading the description, the participants were presented with 24 items and were asked to rate on a 1–6 scale how sure they were that the item had been a part of the description of the client; the category labels (as coded here) were 1 = *I am fairly positive that it was not in the paragraph I read*, 2 = *I am fairly sure that it was not ...*, 3 = *I am undecided but I think that it was not ...*, 4 = *I am undecided but I think that it was ...*, 5 = *I am fairly sure that it was ...*, and 6 = *I am fairly positive that it was ...*. For the present analysis, categories 1 and 2 and categories 5 and 6 were combined, giving a 1–4 scale. Twelve of the items were old (i.e., a part of the description of the client), and 12 were new. Of the 12 old items, 4 were considered to be neutral traits (e.g., "Rob describes himself as energetic"), and 8 were considered to be black stereotypes (e.g., "Rob has a brother who is a gang member"). Similarly, for the 12 new items, 4 were neutral, and 8 were black stereotypes.

A basic question of interest in the research was whether telling the participants that the client was white or black would affect their memory (i.e., detection) for old stereotyped items or would bias their responses in some way. Because of the small sample size per participant ($N = 24$), pooled data are analyzed here; of course, the analysis can also be performed on individual data, given a large enough sample size. The data, as set up in the SPSS file, are shown in Table 4; note that $N = 431$, instead of $18 \times 24 = 432$, because 1 participant skipped the 24th trial.

A first step is to fit unequal variance SDT models separately to the data from the two groups of participants (i.e.,

told that Rob was white or black) and to compare the parameter estimates across the groups. This can be done by using **split file** under the data options and choosing the dummy coded race variable (X_1) as a grouping variable. For the model considered here, however, the spacing of the response criteria was restricted to be equal across the groups, because there was no theoretical reason to suppose that the criteria spacing would be affected by the stated race of the client (results from a separate group analysis were also consistent with this). Although the spacing of the response criteria were restricted to be equal across the groups, the location of the criteria was allowed to differ (by a constant) across the groups, which reflects a type of response bias. This model was compared with a model with additional restrictions, as will be described below.

To start, note that from the perspective of SDT, there are four underlying distributions that correspond to the four classes of items: new–neutral items, new–stereotyped items, old–neutral items, and old–stereotyped items, which will be referred to as NN, NS, ON, and OS items, respectively. Three dummy coded variables—say, X_2, X_3, and X_4—are used to indicate the NS, ON, and OS items, respectively, with NN items serving as the reference (and so the NN distribution has a mean of zero and a variance of unity, for the probit link). The variables X_2, X_3, and X_4 are included as main effects in both the location and the scale parts of the model; including them in the location part allows the NS, ON, and OS distributions to have a different location than does the NN distribution, whereas including them in the scale part allows the variances of the distributions to differ from unity. Let X_1 be a variable with a value of 0 for participants who were told that the client was white and of 1 for participants who were told that the client was black. To allow the locations and variances of the NS, ON, and OS distributions to differ across the groups, interaction terms with X_1 are created—specifically, X_1*X_2, X_1*X_3, and X_1*X_4—and are included in the location and scale parts of the model. Finally, including X_1 as a main effect in the location part of the model allows for response bias, in that it allows the location of the response criteria to differ by a constant across the groups, while maintaining the same spacing.

The model just described can be written as Equation 8 (see bottom of the page), where X is a vector consisting of the regressors shown on the right side of the equation, b is the response bias parameter, d_i are the detection parameters for distributions 2 (NS), 3 (ON), and 4 (OS), Δd_i is the change in the location of distribution i for the participants who were told that the client was black, a_i are scale parameters, and Δa_i is the change in the scale parameter for distribution i. Note that one could also allow the variance of NN items to differ across the groups by including X_1 in the denominator, but this was not done because there was

Table 4
Data as Organized for SPSS: Gushue (2002) Pilot Data

X_1	X_2	X_3	X_4	Response	Frequency
0	0	0	0	1	16
0	0	0	0	2	10
0	0	0	0	3	5
0	0	0	0	4	5
0	1	0	0	1	43
0	1	0	0	2	12
0	1	0	0	3	8
0	1	0	0	4	9
0	0	1	0	1	4
0	0	1	0	2	6
0	0	1	0	3	7
0	0	1	0	4	19
0	0	0	1	1	3
0	0	0	1	2	7
0	0	0	1	3	5
0	0	0	1	4	56
1	0	0	0	1	24
1	0	0	0	2	6
1	0	0	0	3	3
1	0	0	0	4	3
1	1	0	0	1	50
1	1	0	0	2	10
1	1	0	0	3	2
1	1	0	0	4	10
1	0	1	0	1	9
1	0	1	0	2	5
1	0	1	0	3	4
1	0	1	0	4	18
1	0	0	1	1	7
1	0	0	1	2	5
1	0	0	1	3	1
1	0	0	1	4	59

Notes—$X_1 = 1$ for participants who were told that the client was black (and 0 for white), $X_2 = 1$ indicates a new–stereotyped item, $X_3 = 1$ indicates an old–neutral item, and $X_4 = 1$ indicates an old–stereotyped item.

no reason to expect that telling the participants the race of the client would affect the variance of NN items.

The SPSS syntax for the model is

```
PLUM
y WITH x1 x2 x3 x4
/CRITERIA = CIN(95) DELTA(0) LCONVERGE(0)
MXITER(100) MXSTEP(5)
PCONVERGE(1.0E-6) SINGULAR(1.0E-8)
/LINK = PROBIT
/LOCATION = x1 x2 x3 x4 x1*x2 x1*x3 x1*x4
/PRINT = FIT KERNEL PARAMETER SUMMARY
/SCALE = x2 x3 x4 x1*x2 x1*x3 x1*x4
```

The first row of Table 5 shows the results. As a first step, it is useful to check the df given in the SPSS output. In this case, the data being analyzed consist of an 8×4 frequency table, with the eight rows corresponding to the four types of items, for each of the two stated races, and the four columns corresponding to the 1–4 confidence rat-

$$\Phi^{-1}\left[p\left(Y \leq k | X\right)\right] = \frac{c_k - bX_1 - d_2X_2 - d_3X_3 - d_4X_4 - \Delta d_2 X_1X_2 - \Delta d_3 X_1X_3 - \Delta d_4 X_1X_4}{\exp\left(a_2X_2 + a_3X_3 + a_4X_4 + \Delta a_2 X_1X_2 + \Delta a_3 X_1X_3 + \Delta a_4 X_1X_4\right)}, \tag{8}$$

Table 5
Goodness-of-Fit Statistics for Two Models: Gushue (2002) Pilot Data

Model	LR	df	p	X^2	df	p	$-2LL$	Difference
Equation 8	5.64	8	.69	5.36	8	.72	826.549	
Restricted model (see text)	9.54	12	.66	9.33	12	.67	830.450	3.901

Notes—LR, likelihood ratio goodness-of-fit statistic; X^2, Pearson goodness-of-fit statistic; $-2LL$, minus two times the estimated log likelihood.

ing response. For each row, only three of the four frequencies are free to vary, because the row totals are fixed by design (i.e., fixed by the number of presentations of each type of item for each group), and so the fourth frequency is not free. Thus, there are $8 \times 3 = 24$ observations. With respect to the number of parameters, there are three response criteria and 13 additional parameters in the full model, as is shown by Equation 8, giving a total of 16 parameters. Thus, the df are $24 - 16 = 8$, as is shown in the first row of Table 5. The LR and chi-square goodness-of-fit statistics in the first row of Table 5 show that the model describes the data, in that neither statistic is significant.

The second row of Table 5 shows results for a model where detection and the variance were restricted to be equal across the groups for NN, NS, and ON items, whereas detection and the variance for OS items were allowed to differ across the groups. Thus, the restricted model sets $\Delta d_2 = \Delta d_3 = \Delta a_2 = \Delta a_3 = 0$. This model is nested within the first model considered above, and the difference in the $-2LL$ across the models provides a likelihood ratio test of the restricted model. The second row of Table 5 shows that the difference is 3.90, which with $12 - 8 = 4$ df gives $p = .42$, and so the restricted model is not rejected.

Table 6 shows the parameter estimates, standard errors, p values, and 95% confidence intervals for a fit of the restricted model. As was noted above, a basic question of interest was whether telling the participants that the client was white or black would affect their memory for OS items or bias their responses in some way. With respect to response bias, the first row of Table 6 shows that the estimate of the response bias parameter b is -0.37, which is significantly different than zero at the .05 level. A negative value of b indicates that the response criteria for the participants who were told that the client was black were to the right (by 0.37) of the criteria for those who were told that the client was white, with the spacing equal across the groups (i.e., the distributions are shifted to the left, or equivalently, the criteria are shifted to the right by a constant). Thus, the results suggest that being told that the client was black led to response bias, in that the participants who were told that the client was black were more conservative with respect to reporting that they had seen an item earlier, as compared with the participants who were told that the client was white.

With respect to possible differences in memory for OS items across the groups, the estimates of Δd_4 and Δa_4 are relevant. Table 6 shows that the estimate of Δd_4 is quite large (2.05), which indicates better memory for OS items

for the participants who were told that the client was black, but the SE (and 95% CI) is quite large and the estimate is not significantly different from zero; this is also the case for Δa_4. Thus, the pilot study does not provide evidence for an effect of stated race on memory for OS items (in terms of a change in either d_4 or a_4), although the large values of the point estimates of Δd_4 and Δa_4 suggest that, perhaps with a larger sample size (and smaller standard errors), an effect might be found.

Table 6 also shows that the estimate of detection of ON items was 1.29, whereas the estimate was 2.05 for OS items. This suggests that OS items might be remembered better than ON items, but the confidence intervals are large and overlap, and so the difference is not significant. It is interesting to note, however, that the point estimates of the detection parameters suggest that the distributions are ordered from left to right as NS, NN, ON, and OS, which is consistent with a result, found in other memory research, referred to as the *mirror effect* (see Glanzer & Adams, 1985). Finally, Table 6 shows that the estimates of $a_2 = \ln \sigma_2$, $a_3 = \ln \sigma_3$, and $a_4 = \ln \sigma_4$ do not differ significantly from zero, which suggests that the underlying normal distributions have standard deviations of unity. I leave it to the reader to explore some other SDT models and parameter restrictions with this data.

CONCLUSION

The availability of general ordinal regression models in the PLUM procedure of SPSS allows researchers to take full advantage of the conceptualization and analysis offered by signal detection theory. Ideally, the present article will lead to a more widespread use of the models.

Table 6
Parameter Estimates for the Restricted Model:
Gushue (2002) Pilot Data

Parameter	Estimate	SE	p	95% CI
b	$-0.372*$	0.180	.039	$-0.725, -0.018$
d_2	-0.432	0.290	.137	$-1.001, 0.137$
d_3	$1.289*$	0.271	$<.001$	$0.759, 1.820$
d_4	$2.049*$	0.500	$<.001$	$1.069, 3.028$
Δd_4	2.117	1.395	.129	$-0.617, 4.851$
a_2	0.397	0.232	.087	$-0.058, 0.852$
a_3	0.209	0.252	.407	$-0.286, 0.704$
a_4	0.223	0.317	.481	$-0.397, 0.844$
Δa_4	0.880	0.472	.062	$-0.044, 1.805$

Notes—all distances are with respect to new-neutral items; d_2 = distance of new–stereotyped items; d_3 = distance of old–neutral items; d_4 = distance of old–stereotyped items; a_i is an estimate of $\ln \sigma_i$ for distribution i. *$p < .05$.

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.

COX, D. R., & SNELL, E. J. (1989). *Analysis of binary data* (2nd ed.). New York: Chapman & Hall.

DAYTON, C. M. (1998). *Latent class scaling analysis*. Thousand Oaks, CA: Sage.

DECARLO, L. T. (1997). On the meaning and use of kurtosis. *Psychological Methods*, **2**, 292-307.

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, ON.

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

DORFMAN, D. D., & ALF, E., JR. (1969). Maximum-likelihood estimation of parameters of signal-detection theory and determination of confidence intervals: Rating method data. *Journal of Mathematical Psychology*, **6**, 487-496.

GESCHEIDER, G. A. (1997). *Psychophysics: The fundamentals* (3rd ed.). Hillsdale, NJ: Erlbaum.

GLANZER, M., & ADAMS, J. K. (1985). The mirror effect in recognition memory. *Memory & Cognition*, **13**, 8-20.

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, Inc.

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

GUSHUE, G. (2002). [Signal detection of neutral items and racial stereotyped items]. Unpublished raw data.

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

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

MCCULLAGH, P., & NELDER, J. A. (1989). *Generalized linear models* (2nd ed.). New York: Chapman & Hall.

MCFADDEN, D. (1974). Conditional logit analysis of qualitative choice behavior. In P. Zarembka (Ed.), *Frontiers of Econometrics* (pp. 105-142). New York: Academic Press.

MENARD, S. (2000). Coefficients of determination for multiple logistic regression analysis. *American Statistician*, **54**, 17-24.

METZ, C. (1998). *Rockit* (0.9B Beta Version). Chicago: University of Chicago, Department of Radiology.

NAGELKERKE, N. J. D. (1991). A note on a general definition of the coefficient of determination. *Biometrika*, **78**, 691-692.

SHEU, C.-F., & HEATHCOTE, A. (2001). A nonlinear regression approach to estimating signal detection models for rating data. *Behavior Research, Methods, Instruments, & Computers*, **33**, 108-114.

STATACORP (2001). *Stata Statistical Software: Release 7*. College Station, TX: Author.

SWETS, J. A. (1996). *Signal detection theory and ROC analysis in psychology and diagnostics: Collected papers*. Hillsdale, NJ: Erlbaum.

SWETS, J. A., TANNER, W. P., JR., & BIRDSALL, T. G. (1961). Decision processes in perception. *Psychological Review*, **68**, 301-340.

TOSTESON, A. N. A., & BEGG, C. B. (1988). A general regression methodology for ROC curve estimation. *Medical Decision Making*, **8**, 204-215.