

Signal Detection Theory and Generalized Linear Models

Lawrence T. DeCarlo
Fordham University

Generalized linear models are a general class of regressionlike models for continuous and categorical response variables. Signal detection models can be formulated as a subclass of generalized linear models, and the result is a rich class of signal detection models based on different underlying distributions. An example is a signal detection model based on the extreme value distribution. The extreme value model is shown to yield unit slope receiver operating characteristic (ROC) curves for several classic data sets that are commonly given as examples of normal or logistic ROC curves with slopes that differ from unity. The result is an additive model with a simple interpretation in terms of a shift in the location of an underlying distribution. The models can also be extended in several ways, such as to recognize response dependencies, to include random coefficients, or to allow for more general underlying probability distributions.

Signal detection theory (SDT) arose as an application of statistical decision theory to engineering problems, in particular, the detection of a signal embedded in noise. The relevance of the theory to psychophysical studies of detection, recognition, and discrimination was recognized early on by Tanner and Swets (1954) and others (see Green & Swets, 1966). SDT has in recent years been applied to a wide variety of research in psychology (see Gescheider, 1997; MacMillan & Creelman, 1991; McNicol, 1972; Swets, 1986) and in other areas, such as medical research, weather forecasting, survey research (see Swets, 1996), and marketing research (Singh & Churchill, 1986).

This article shows that *generalized linear models* (GLMs), which are a general class of regressionlike models for continuous and categorical response variables, provide a useful framework for signal detection theory. In particular, signal detection models can be formulated as a subclass of GLMs, and the result is a

rich class of signal detection models based on different underlying distributions. An example is a signal detection model based on the extreme value distribution. The extreme value model is shown to yield unit slope receiver operating characteristic (ROC) curves for several classic data sets that are commonly given as examples of normal or logistic ROC curves with slopes that differ from unity. This has implications for research and theory in several areas, such as for current memory research.

GLMs also offer a flexible regressionlike approach to signal detection data, and the basic model can be extended in several ways. For example, additional variables can be included, such as covariates one wishes to control for. The possibility of dependence among responses can be examined through the use of time series and longitudinal extensions of GLMs. In the last section I comment on these and other extensions and note directions for future research.

SDT and Logistic Regression

In this section I show that SDT with logistic underlying distributions leads directly to a logistic regression model, which provides a convenient starting point for the generalization to GLMs below. Logistic regression provides a simple way to estimate and test the signal detection parameters and can be used for both binary and rating response data.

SDT

Figure 1 shows the basic ideas and parameters of SDT. One idea is that the effect of the presentation of

Lawrence T. DeCarlo, Department of Psychology, Fordham University.

I thank Michael Kubovy and Richard Perline for helpful comments on earlier versions of this article.

Correspondence concerning this article should be addressed to Lawrence T. DeCarlo, who is now at the Department of Human Development, Teachers College, Columbia University, 525 West 120th Street, New York, New York 10027. Electronic mail may be sent to decarlo@msmailhub.tc.columbia.edu.

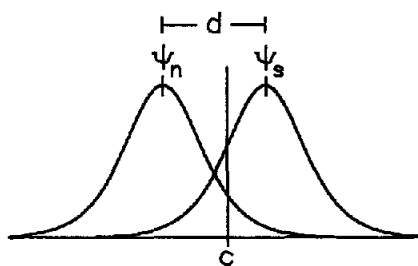


Figure 1. Signal detection theory with logistic underlying distributions and a binary response.

an event, such as a signal or noise, can be represented by an underlying probability distribution, such as the logistic distributions shown in the figure. It is assumed that the distributions associated with signal and noise differ only with respect to location. Note that the underlying distributions can be conceptualized in more than one way, depending on the area of application. For example, they are often thought of in psychophysics as distributions of perception, whereas in memory research they can be viewed as being distributions of familiarity.

A second idea of SDT is that, on each trial, a participant decides whether an event has occurred by using a response criterion. In particular, the participant responds "yes" (reports a signal) if the subjective event (e.g., sensation) falls above the criterion; otherwise he or she responds "no."

As shown in Figure 1, the basic parameters of SDT are c and d . The parameter c is the distance of the response criterion from the mode of the noise distribution (MacMillan & Creelman, 1991, discussed other criterion measures; also see Appendix A of the present article), and the parameter d is a measure of the distance between the two underlying distributions,

$$d = \frac{\psi_s - \psi_n}{\tau},$$

where ψ_s and ψ_n are, respectively, the modes of the signal and noise distributions, and τ is a scale parameter. For logistic distributions, d is simply two times the logarithm of the choice theory parameter α , $d = 2 \log(\alpha)$ (see MacMillan & Creelman, 1991; McNicol, 1972), and τ is related to the standard deviation σ by $\sigma = \tau\pi/\sqrt{3}$. Multiplying d by $\sqrt{3}/\pi$ gives the distance in standard deviations, which is the effect size measure δ used in meta-analysis (see Hasselblad & Hedges, 1995). For normal distributions, $\sigma = \tau$ and d is the traditional measure d' . SDT models based on these and other distributions are considered here, so I

simplified the notation by using d throughout for the distance measure and c for the response criterion, both of which are scaled with respect to τ .

An important aspect of SDT is that the parameters d and c separate sensory (memory, etc.) factors from decision factors, respectively.

The area to the right of the criterion under the signal distribution shown in Figure 1 gives the probability of a response of "yes" when a signal is presented, that is, the probability of a *hit*, which for the logistic distribution is

$$p(Y = 1|S) = \frac{1}{1 + \exp[(c - \psi_s)/\tau]},$$

where $p(Y = 1|S)$ is the conditional probability that the participant responds "yes" (reports a signal) given that a signal was presented, and $(c - \psi_s)/\tau$ is the scaled distance of the criterion from the mode of the signal distribution. The right side of the above is the *survival function* for the logistic distribution; it gives the area to the right of the criterion and is simply 1 minus the cumulative distribution function.

Similarly, the probability that the participant says "yes" given that noise was presented, $p(Y = 1|N)$, which is the probability of a *false alarm*, is given by the area to the right of the criterion under the noise distribution, which is

$$P(Y = 1|N) = \frac{1}{1 + \exp[(c - \psi_n)/\tau]}.$$

Next, consider the log odds or *logit* transform, which is

$$\text{logit}(p) = \log\left(\frac{p}{1-p}\right),$$

where \log is the natural logarithm. Applying the logit transform to the hit and false alarm probabilities shown above gives

$$\begin{aligned} \text{logit } p(Y = 1|S) &= \frac{\psi_s - c}{\tau} \\ \text{logit } p(Y = 1|N) &= \frac{\psi_n - c}{\tau}. \end{aligned} \quad (1)$$

The above logits represent, respectively, the log odds that the participant says "yes" to a signal and the log odds that he or she says "yes" to noise. Equation 1 shows that the logit transformed probabilities have a simple relationship to the theoretical parameters. For example, one can easily solve for the parameters d and c : the difference between the logit hits and logit

false alarms gives d , and -1 times the logit false alarms gives c ; substituting observed proportions for the probabilities gives estimates of the parameters (see Appendix A; also, $-1/2$ times the sum of the two logits gives another criterion measure, c').

The next section shows that the above two logits are brought together in a logistic regression model.

Logistic Regression

Logistic regression is a generalization of ordinary linear regression to situations where the dependent variable is binary or polytomous. Textbook introductions have been provided by Hosmer and Lemeshow (1989) and Kleinbaum (1994); Strauss (1992) provided an informative article.

The two components of Equation 1 can be combined in a logistic regression model as

$$\text{logit } p(Y = 1|X) = \frac{\psi_n - c}{\tau} + \frac{\psi_s - \psi_n}{\tau} X,$$

where X is coded as 1 for signal and 0 for noise (dummy coding). The reader can verify that when $X = 1$, the above gives the log odds (logit) of a hit, as shown in Equation 1, and when $X = 0$, it gives the log odds of a false alarm. Given the definition of d shown above, and setting $\psi_n = 0$ and $\tau = 1$, the model can be written more simply as

$$\text{logit } p(Y = 1|X) = -c + dX. \quad (2)$$

Equation 2 shows that the slope and intercept of the logistic regression model have a simple yet important relation to the signal detection parameters: The coefficient of X (the slope) gives the distance measure d , and the intercept gives $-c$. Logistic regression, therefore, can be used to estimate and test the signal detection parameters; it brings the utility of regression and analysis of variance (ANOVA) to signal detection research.

Rating Responses

Equation 2 can be extended to rating response experiments by modeling cumulative probabilities; McCullagh (1980) noted advantages of this approach, whereas Agresti (1989) provided a tutorial on this and other approaches to modeling ordinal categorical data.

Participants in rating response signal detection experiments are typically asked to use a rating scale to indicate how confident they are that a signal has been presented. Figure 2 shows the underlying theory for four response categories, say from 1 = *sure noise* to 4 = *sure signal*. It is assumed that participants choose

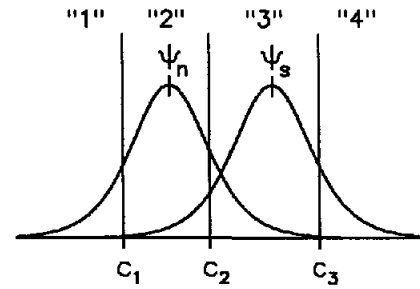


Figure 2. Signal detection theory with a rating response.

among the response categories by using several criteria, one fewer than the number of response categories. As before, SDT focuses on the area to the right of each criterion, which for the logistic distribution is

$$P(Y > j|S) = \frac{1}{1 + \exp[(c_j - \psi_s)/\tau]}$$

$$P(Y > j|N) = \frac{1}{1 + \exp[(c_j - \psi_n)/\tau]}, \quad (3)$$

where $1 \leq j \leq k - 1$, k is the number of response categories, c_j are the multiple criteria, and the function on the right is the survival function for the logistic distribution.

An ordinal regression model. Using the logit transform, Equation 3 can be written as an ordinal regression model,

$$\text{logit } p(Y > j|X) = -c_j + dX,$$

where $\psi_n = 0$, $\tau = 1$, and c_j are the distances of the criteria from the mode of the noise distribution. Note that the above simply replaces $p(Y = 1|X)$ of Equation 2 with $p(Y > j|X)$; when $k = 2$, the model reduces to Equation 2 with responses of "no" and "yes" indicated by "1" and "2," respectively.

With respect to fitting the model, although the approach in SDT (and survival analysis) is to model the area to the right of the criterion, which is given by the survival function, the approach for GLMs is to model the area to the left of the criterion, which is given by the cumulative distribution function (CDF). Using the relation $\text{logit}(p) = -\text{logit}(1 - p)$, the model written in terms of cumulative probabilities is

$$\text{logit } p(Y \leq j|X) = c_j - dX, \quad (4)$$

where the above logits are *cumulative logits*. Note that the signs of the parameters are simply reversed (this is not the case, however, for asymmetric distributions).

Equation 4 is a logistic signal detection model for rating response experiments; it shows that the coefficient of X , multiplied by -1 , gives an estimate of the distance measure d , and the intercepts give the multiple criteria c_j . The use of cumulative logits is one way to take the ordinal nature of the response into account in the analysis.

Logistic ROC curves and proportional odds. An appealing aspect of rating response experiments is that an ROC curve can be obtained for data from a single session (for binary response experiments, multiple sessions or conditions are needed to obtain an ROC curve). ROC curves show the relation between the hit and false alarm probabilities for varying criteria. ROC curves are of interest in SDT because, when linearized, a slope of unity implies that the variances of the underlying distributions are equal. Thus, ROC curves are often used in signal detection research to assess the equal variance assumption.

The equation for the linearized ROC curve can be derived from the two components of Equation 3 by taking logits, subtracting, and rearranging terms, which gives

$$\text{logit } p(Y > j|S) = d + \text{logit } p(Y > j|N).$$

The above shows that the intercept of the linearized ROC curve is d and the slope is unity. A plot of the logit transformed proportions of hits versus false alarms gives an empirical ROC curve.

Subtracting $\text{logit } p(Y > j|N)$ from each side of the above shows that the difference between the cumulative logits is simply d . Exponentiating the difference gives

$$\frac{p(Y > j|S)/[1 - p(Y > j|S)]}{p(Y > j|N)/[1 - p(Y > j|N)]} = \frac{\text{odds } p(Y > j|S)}{\text{odds } p(Y > j|N)} = \exp(d).$$

The above shows that, if the ROC curve has a slope of unity, then the odds are proportional by a factor of $\exp(d)$, independent of the criteria; a test of this *proportional odds* assumption is frequently included in software for logistic regression. With respect to SDT, the test of proportional odds is equivalent to a test of the assumption that the ROC curve has a slope of unity; the test provides a useful supplement to visual inspection of the ROC curve.

With respect to Equation 4, the test of proportional odds is a test of whether d (the log odds ratio) is constant across the response categories. The test com-

pares the fit of two models, one with a single value of d ,

$$\text{logit } p(Y \leq j|X) = c_j - dX,$$

against one where d is allowed to vary across the response categories (denoted d_j),

$$\text{logit } p(Y \leq j|X) = c_j - d_j X,$$

with $1 \leq j \leq k-1$. A comparison of the fit of the two models provides a test of the null hypothesis of equal slopes, $H_0: d_1 = d_2 = \dots = d_{k-1}$. Specifically, one can perform a *likelihood ratio* (LR) test (see Appendix A) by subtracting -2 times the maximized log likelihood for the second model (with different d_j) from that for the first model (with constant d). The degrees of freedom are the difference in the number of parameters across the two models, which is $df = k - 2$. The examples in Appendixes A and B illustrate the analysis for binary and rating response experiments with SAS (SAS Institute, 1989) and SPSS (SPSS, 1994) programs.

In sum, the approach to signal detection through logistic regression offers advantages: It offers a model-based approach to SDT that can be used to estimate and test the parameters, to test hypotheses about the ROC curve, and to answer basic questions of interest in research, such as whether the parameters differ across conditions or groups, for example. It also helps to show that a basic goal is to find a simple model that provides a quantitative summary of the data.

In the next section I show that GLMs generalize the above approach and offer a more general class of signal detection models. GLMs maintain the advantages of logistic regression yet offer increased flexibility.

SDT and GLMs

A general introduction to GLMs is beyond the scope of this article; Agresti (1996), Dobson (1990), and Liao (1994) provided introductions, McCullagh and Nelder (1989) offered an in-depth presentation, and Hilbe (1994) provided a review of software for GLMs. Rather, the focus here is on aspects of GLMs that are particularly relevant to SDT. In particular, it is shown that signal detection models based on different underlying distributions can easily be considered by using GLMs with different "link" functions.

To begin, note that, among other things, it is assumed in classical regression that the mean response

is linearly related to the predictors and also that the response variable is normally distributed (for confidence intervals and hypothesis tests). GLMs, as introduced by Nelder and Wedderburn (1972), generalize classical regression models with respect to both of these aspects. For example, GLMs require only that a transformation of the mean response be linearly related to the predictors, and they offer a more general class of probability distributions for the response variable (both continuous and discrete, so GLMs unify methods for continuous and categorical data). GLMs also relax the assumption of constant variance for the response and require only that the variance be a known function of the mean (possibly with an additional dispersion parameter).

The above generalizations are achieved by viewing the model as consisting of several components: a random component, a systematic component, and a link function (McCullagh & Nelder, 1989). The random component is concerned with the probability distribution for the response variable. As noted above, GLMs are flexible in this regard, in that they allow one to choose distributions from the exponential family, which includes continuous and discrete distributions, such as the normal, gamma, binomial, and Poisson distributions (quasi-likelihood offers a further extension; see McCullagh & Nelder, 1989). For example, the binomial distribution is used for the binary response models discussed above. The systematic component is a linear predictor $\eta = \Sigma \beta X$, as is used in ordinary regression models. For signal detection models, the basic linear predictor is

$$\eta = c_j - dX,$$

where X indicates signal or noise, c_j are one or more criteria, and d is the scaled distance measure.

The focus here is on the *link* function, which specifies how the random and systematic components are related. In particular, the link function is a monotonic differentiable function $g(\cdot)$ that "links" the mean response, which for SDT is the response probability p , to the linear predictor η ,

$$g(p) = \eta.$$

For example, for the logit link $g(p)$ is logit p , and the model is

$$\log \left(\frac{p}{1-p} \right) = \eta.$$

From the perspective of SDT, an important aspect of the link function is that its inverse, g^{-1} , corresponds to

a cumulative distribution function for the underlying distributions. For example, the inverse of the model with logit link is

$$p = \frac{1}{1 + \exp(-\eta)},$$

and the term on the right is the CDF for the logistic distribution. Combined with linear predictor $\eta = c_j - dX$, the above gives a signal detection model based on the logistic distribution.

It follows that signal detection models based on different underlying distributions can be obtained by using different link functions. For example, the inverse normal (probit) link gives the traditional signal detection model based on the normal distribution, in that the inverse of the link function gives the CDF for the standard normal. In this case, d is the traditional distance measure d' . Note that a fit of the normal theory signal detection model generally gives results similar to those obtained with the logistic model, with the exception that the parameters are scaled differently. For example, Agresti (1990) noted (pp. 103–104) that estimates for logistic models tend to be about 1.6–1.8 times larger than those for probit (normal) models (partly because of the different scaling, that is, the standard deviation for the normal is $\sigma = \tau$ but is $\sigma = \tau\pi/\sqrt{3} = 1.8\tau$ for the logistic). Similarly, 1.7 is used as a scale factor in item response theory because it minimizes the maximum difference between the logistic and normal distributions (see Camilli, 1994). When comparing logistic and normal models, therefore, the arbitrary scaling of the parameters should be kept in mind.

Another widely used link for GLMs is the complementary log-log link

$$\log[-\log(1-p)] = \eta,$$

the inverse of which is

$$p = 1 - \exp[-\exp(\eta)],$$

and the term on the right in this case is the CDF for an extreme value distribution. This model has important implications for SDT and is discussed in the next section.

It should be apparent that the above models all have a similar structure, in that a transformation of the response probability has a linear relationship to the signal detection parameters. A general signal detection model is

$$p(Y \leq j|X) = F(c_j - dX),$$

where F is a CDF for the underlying distribution. The inverse of the above gives a generalized linear model for SDT, which is

$$g[p(Y \leq j|X)] = c_j - dX.$$

Signal detection models, in other words, are a subclass of GLMs where the inverse link function g^{-1} corresponds to a cumulative distribution function F for the underlying variable. Because X is (usually) simply dichotomous (signal or noise), signal detection models are a form of generalized ANOVA models.

The above generalization encompasses a variety of signal detection models. For example, in addition to the logistic, normal, and extreme value models considered here, signal detection models based on exponential, uniform (the identity link), Weibull, Rayleigh, Cauchy, Laplace, Pareto, and other distributions can be formulated. Thus, GLMs provide a unified framework for signal detection models, some of which have previously been considered by Egan (1975), Green and Swets (1966), Morgan (1976), and Swets (1986).

It follows that different links can be used to see if an alternative distribution gives a unit slope ROC curve. If it does, then the link offers a monotonic transformation to additivity, in which case the model, analysis, and interpretation of the results are simplified; it is analogous to finding simple main effects with no interaction in ANOVA. In the next section I illustrate the approach with a signal detection model based on the extreme value distribution.

SDT With Extreme Value Distributions

Extreme value distributions differ from logistic and normal distributions in that they are asymmetric. They are widely used in statistics for GLMs (see Agresti, 1990; Dobson, 1990; Liao, 1994; McCullagh & Nelder, 1989) and in survival and reliability analysis (e.g., Klein & Moeschberger, 1997; Lawless, 1982); Johnson, Kotz, and Balakrishnan (1995) provided background and extensive references. The extreme value distribution considered here is the distribution of *smallest extremes*, which is negatively skewed; it arises as the limiting distribution of smallest values of independent and identically distributed random variables and in general is useful for describing the behavior of a system made up of a large number of components operating in parallel. The normal distribution is often motivated by the central limit theorem and the averaging of events, whereas the extreme value distribution can be motivated by parallel pro-

cessing and extreme (minima or maxima) events (cf. Wandell & Luce, 1978).

Figure 3 shows the theory for binary responses and extreme value distributions. As before, SDT is concerned with the probabilities of hits and false alarms, which are given by the area to the right of the criterion under the signal and noise distributions,

$$p(Y = 1|S) = \exp \left[-\exp \left(\frac{c - \psi_s}{\tau} \right) \right]$$

$$p(Y = 1|N) = \exp \left[-\exp \left(\frac{c - \psi_n}{\tau} \right) \right],$$

where ψ_s and ψ_n are the modes of the signal and noise distributions, as shown in Figure 3; τ is a scale parameter related to the standard deviation by $\sigma = \tau\pi/\sqrt{6}$; and the term on the right is the survival function for the distribution of smallest extremes. Taking one minus the survival function gives the cumulative distribution function, shown above as the inverse of the complementary log-log link.

Extreme value ROC curves. The linearizing transform in this case is the log-log transform, $-\log[-\log(p)]$, which applied to the above gives

$$-\log[-\log(p[Y = 1|S])] = \frac{\psi_s - c}{\tau}$$

$$-\log[-\log(p[Y = 1|N])] = \frac{\psi_n - c}{\tau}.$$

Subtracting and rearranging terms gives a linearized ROC curve, which is

$$-\log[-\log(p[Y = 1|S])] = d + [-\log(-\log(p[Y = 1|N]))],$$

where d is the distance between the modes of the two distributions (and the means, which are at $\psi - 0.5772\tau$; 0.5772 is Euler's constant γ). The equation

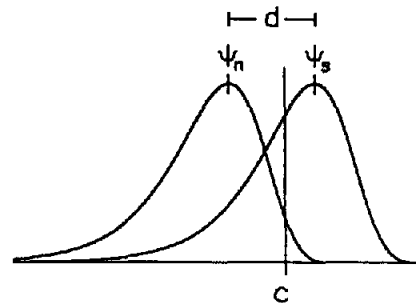


Figure 3. Signal detection theory with extreme value underlying distributions and a binary response.

for the linearized ROC curve shows that the equal variance assumption can be assessed for the extreme value model by using an ROC plot with log-log coordinates; the result should be a straight line with a slope of unity. For rating response experiments, $p(Y = 1|X)$ is replaced by $p(Y > j|X)$.

Subtracting the log-log transformed false alarm probability from each side of the above and exponentiating shows that the ratio of the log hit and log false alarm probabilities is a constant, $\exp(-d)$, that depends only on the distance between the two underlying distributions and not on c_j . This property, known as *proportional hazards* (because the cumulative hazard functions are proportional), is similar to the proportional odds property discussed earlier and is equivalent to an extreme value ROC curve with a slope of unity, or equal d_j in terms of the extreme value model.

Because normal ROC plots have been widely used in signal detection research, it is of interest to see what extreme value ROC curves look like when plotted on inverse normal coordinates. The upper panel of Figure 4 shows three ROC curves (and a diagonal) corresponding to an extreme value SDT model with equal variances; the three curves are for different values of d (2, 4, and 6). The panel shows that, if the underlying distributions are extreme value distributions with equal variances, as shown in Figure 3, then the ROC curves on log-log coordinates are linear with slopes of unity. The lower panel shows the curves on inverse normal coordinates. In this case, the ROC curves are bowed, but near linear, with slopes less than unity (and a pattern of decreasing slopes for increasing d ; a similar pattern has been noted for exponential and other distributions by, e.g., Swets, 1973). The figure shows that nonunit slope normal ROC curves might arise because the underlying distributions are skewed.

To illustrate, Figure 5 shows the results for an influential recognition memory experiment of Egan (1958). Sixteen participants used a 7-category rating response to indicate their confidence that a word had been presented earlier in the experiment. The top panel is similar to Figure 19 of Egan (1958; the data in Figure 5 were estimated from Egan's figure) and shows data averaged over four sets of 4 participants (the quartiles in terms of d'). The three ROC curves all have slopes less than unity (and the slopes decrease as d increases, which is the same pattern shown in Figure 4). This has been interpreted as showing that the variance of the old word distribution is larger than

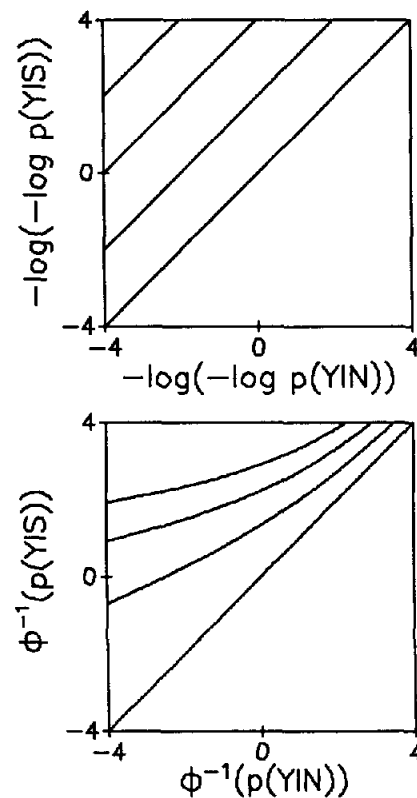


Figure 4. Extreme value receiver operating characteristic curves on log-log coordinates (upper panel) and inverse normal coordinates (lower panel).

the variance of the new word distribution, which in turn has influenced theories of recognition memory.

The lower panel of Figure 5 shows the data on log-log coordinates. In this case, the three ROC curves all have slopes close to unity. Thus, the extreme value model offers an important simplification, in that it is consistent with the equal variance assumption and results in a simple model. It also has implications about the interpretation of the data. For example, Swets's (1986) question of "Or why do data show 'old' words to be more variable than 'new' words (slopes near 0.70)?" (p. 195) becomes a question about why the underlying distributions are skewed.

A second example is a widely analyzed binary response light detection experiment of Swets, Tanner, and Birdsall (1961). Figure 6 shows the results for 4 participants who took part in 12–13 sessions (some sessions are not shown because of zero proportions); the data (proportions) are presented in Dorfman and Alf (1968; also in Table 6.2 of Coombs, Dawes, & Tversky, 1970). When plotted on inverse normal co-

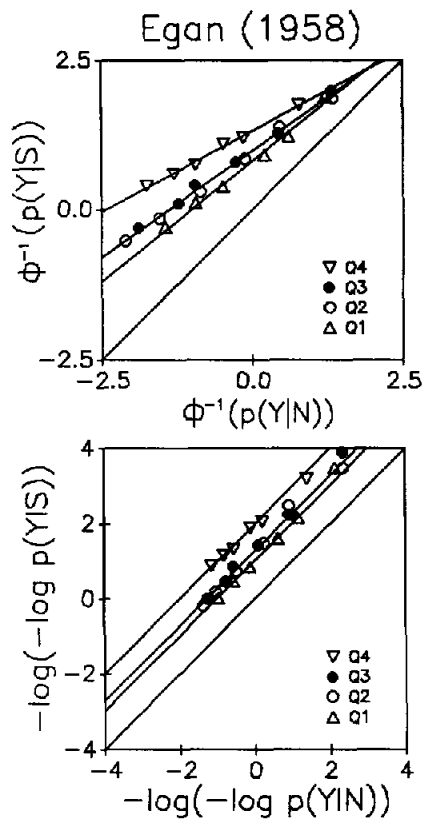


Figure 5. ROC plots for a rating response word recognition experiment of Egan (1958). Q1–Q4 represent the first to fourth quartiles; $p(Y|N)$ is the probability of reporting a signal (Y) given that noise (N) was presented (a false alarm); $p(Y|S)$ is the probability of reporting a signal given that a signal (S) was presented (a hit). The upper panel is from Egan's Figure 19 and shows the average results for four sets of 4 participants each. The lower panel shows the same data on log-log coordinates.

ordinates, the ROC curves are linear with slopes less than unity, as shown by Swets et al. (1961, Figure 14) and others (Dorfman & Alf, 1968, Figure 1). This has been interpreted as showing that the underlying distributions have unequal variances. Figure 6, on the other hand, shows that on log-log coordinates the ROC curves are linear with slopes close to unity for all 4 participants, which implies that the underlying distributions are extreme value distributions with equal variances. Again, this has implications for the interpretation of the results.

It is interesting to note that a recent textbook example of nonunit slope ROC curves (MacMillan & Creelman, 1991, Table 3.5, p. 65) is also a "textbook" example of unit slope extreme value ROC curves.

An extreme value signal detection model. The transformed hit and false alarm probabilities shown above can be combined in a regression model as

$$-\log[-\log(p\{Y = 1|X\})] = -c + dX,$$

where c is the distance of the criterion from the mode of the noise distribution, and d is the distance between the two underlying distributions. Note that the above is simply Equation 2 with a log-log link.

With respect to fitting the model, it should be kept in mind that the approach for GLMs is to model the CDFs. In this case, this can be done by using the complementary log-log link, $\log[-\log(1 - p)]$, which is a standard link in software for GLMs, and by modeling the conditional probability of a response of "no"

$$\log[-\log(1 - p\{Y = 0|X\})] = c - dX,$$

where $p\{Y = 0|X\}$ is the probability that the participant responds "no" to signal or noise. The model can be applied to rating response experiments by replacing the above probabilities with cumulative probabilities, which gives

$$\log[-\log(1 - p\{Y \leq j|X\})] = c_j - dX, \quad (5)$$

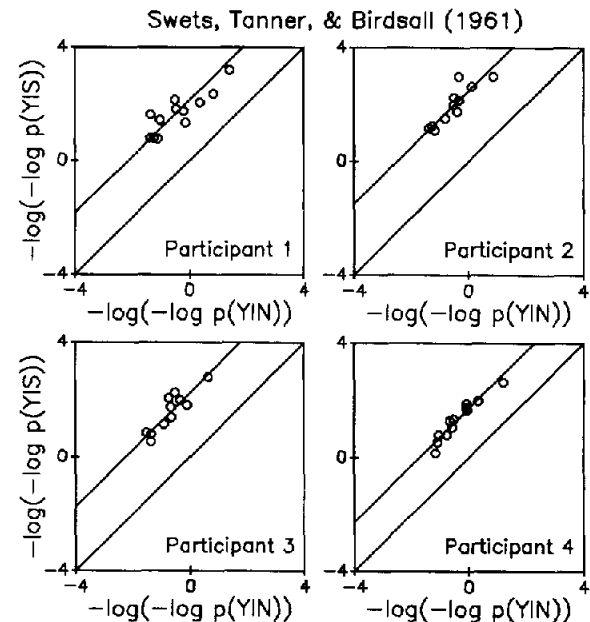


Figure 6. Extreme value receiver operating characteristic plots for 4 participants from a binary response light detection experiment of Swets et al. (1961); $p(Y|N)$ is the probability of reporting a signal (Y) given that noise (N) was presented (a false alarm); $p(Y|S)$ is the probability of reporting a signal given that a signal (S) was presented (a hit).

where $1 \leq j \leq k-1$, k is the number of response categories, and the responses are numbered from left to right (sure noise to sure signal) as 1, 2, 3, and so on. Note that Equation 5 differs from Equation 4 only with respect to the link function. Equation 5 is a proportional hazards model (McCullagh, 1980), because the ratio of the hazard functions is constant, just as the ratio of odds is constant for the logistic (proportional odds) model.

An example of fitting and testing the extreme value signal detection model follows. The results are compared with those obtained for the logistic signal detection model.

Logistic and Extreme Value SDT Models Compared

Green and Swets (1966, p. 102) provided the raw data for a participant from a widely analyzed rating response light detection experiment of Swets et al. (1961; Subject 1, who was also in the binary response experiment of Figure 6). Appendix B provides the data and a SAS program.

Figure 7 shows ROC curves for both the logistic and extreme value models. With logit coordinates, the ROC curve has a slope less than unity (like the curve on inverse normal coordinates, not shown), which has been interpreted as showing that the variance of the signal distribution is larger than that of the noise distribution. On the other hand, with log-log coordinates, the ROC curve has a slope close to unity, which implies that the (extreme value) distributions have equal variance.

Table 1 shows, for fits of the logistic and extreme value signal detection models, the estimated values of d , a score test of equal d_j (proportional odds or hazards), and goodness-of-fit statistics. The maximum likelihood estimate of d is 2.15 for the logistic model and 1.39 for the extreme value model; the larger value for the logistic distribution occurs because of the different scaling (the standard deviations differ by a factor of $\sqrt{2}$) and shape of the two distributions.

The assumption of a unit slope ROC curve can be formally assessed in a manner similar to that used to test for proportional odds. In particular, a likelihood ratio test of equal d_j can be performed (using the $-2 \log$ likelihood) by comparing, for a given link, the fit of

$$g[p(Y \leq j|X)] = c_j - dX$$

to the fit of

$$g[p(Y \leq j|X)] = c_j - d_j X,$$

Swets, Tanner, & Birdsall (1961)

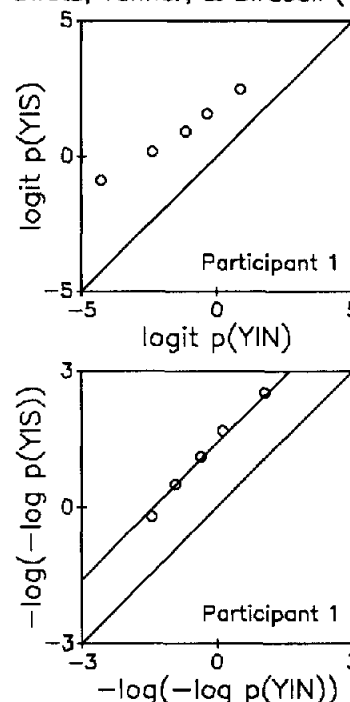


Figure 7. Logistic (upper panel) and extreme value (lower panel) receiver operating characteristic plots for participant 1 from a rating response light detection experiment of Swets et al. (1961); $p(Y|N)$ is the probability of reporting a signal (Y) given that noise (N) was presented (a false alarm); $p(Y|S)$ is the probability of reporting a signal given that a signal (S) was presented (a hit).

to test $H_0: d_1 = d_2 = \dots = d_{k-1}$ with $k-2$ df. From the perspective of SDT, the test of equal d_j can be used to test for unit slope ROC curves for logistic, normal, extreme value, and other models.

Table 1 shows that the null hypothesis of equal d_j is rejected for the logistic model (using the score statistic that SAS provides by default; SAS Institute, 1989) but is not rejected for the extreme value model. Similarly, the goodness-of-fit statistics are close to their df for the extreme value model (indicating good fit) but not for the logistic model. Note that the tests of goodness of fit and equal d_j are equivalent in this case, because both tests assess the fit of the model with unit slope ROC curve against a saturated model.

Figure 7 and Table 1 show that the assumption of a unit slope ROC curve is tenable for the extreme value model, so the results can be summarized by the simple distance measure d . The logistic model, on the other hand, gives an ROC curve with a slope less than unity.

To summarize, a basic aspect of GLMs is that they

Table 1
Results for Swets et al.'s (1961) Rating Response Light Detection
Experiment: Participant 1

	Logistic model		Extreme value model	
Parameter	Estimate	SE	Estimate	SE
d	2.147	.118	1.388	.072
c_1	-0.783	.085	-1.071	.069
c_2	0.366	.081	-0.166	.051
c_3	1.108	.088	0.342	.047
c_4	2.002	.101	0.907	.049
c_5	3.147	.121	1.583	.061
	$-2 \log L = 3872.743$		$-2 \log L = 3845.613$	
Equal-slopes test				
Score χ^2	29.42		5.36	
p value	.000		.252	
df	4		4	
Goodness-of-fit tests				
LR χ^2	32.26		5.12	
p value	.000		.275	
df	4		4	
Pearson χ^2	29.49		5.57	
p value	.000		.233	
df	4		4	

Note. SE = standard error; $-2 \log L = -2 \log$ likelihood.

offer a choice of link functions. With respect to SDT, the inverse of the link function corresponds to a CDF for an underlying random variable. Thus, the approach via GLMs allows one to consider SDT models based on different underlying distributions by using different links.

The above examples show that finding an ROC curve with a slope other than unity on given coordinates has more than one possible interpretation. A common conclusion has been that the variances are not equal across the normal or logistic signal and noise distributions. However, another possibility is that the variances are equal but that the underlying distributions are not normal or logistic.

It should be recognized that finding a unit slope ROC curve on alternative coordinates is equivalent to finding a transformation to additivity, and this in turn simplifies the model, its estimation, and the interpretation of the results. The unit slopes shown in Figures 5, 6, and 7, for example, can be interpreted as showing that presentation of a signal results in a simple shift in the location of an underlying extreme value distribution. This may also have theoretical implications: As noted above, the extreme value distribution arises from parallel processing and extreme events (minima or maxima), and this could suggest something about

underlying processes in word recognition (Figure 5) or light detection (Figure 6).

The approach is also similar to that of Box and Cox (1964), with the difference being that the Box-Cox approach is used in regression analysis, for example, to transform the data (to obtain additivity, constant variance, or normality), whereas the data are not transformed in GLMs; GLMs accomplish the transformation (statistically) through the link function (GLMs transform the fitted values, not the data). McCullagh and Nelder (1989) noted that an advantage of GLMs is that the link can be used to obtain additivity, separately from considerations about the random component of the model, such as constant variance or normality of the response variable (and GLMs offer a wider class of probability distributions for the response variable).

Although there are clearly situations in which alternative coordinates give unit-slope ROC curves, as shown above, this may not be so in some cases. One response is to consider generalizations of the basic theory. For example, SDT has been generalized to allow the variances to differ, as shown below. The resulting model, however, is more complex (e.g., it is nonlinear in the parameters), as is its estimation and interpretation. In addition, there is more than one way

to generalize the theory (see generalized probability distributions below), and different generalizations can lead to different conclusions. This point is relevant to current research and theory on global memory models, for example, because ROC curves have been used to evaluate predictions about the relative variances made by different models (see Ratcliff, Sheu, & Gronlund, 1992). Conclusions about the relative variances, however, depend on the assumed form of the underlying distributions.

Consider, for example, Egan's (1958) rating response experiment on recognition memory for words discussed above. Egan presented a widely reproduced figure (his Figure 20) that shows an ROC curve for the average ratings of 8 participants (the 2nd and 3rd quartiles of Figure 5 of the present article) who saw the word list once before a recognition test, and a second ROC curve for a different group of 8 participants who saw the list twice before the test. The top panel of Figure 8 shows the proportions of hits and false alarms on logit coordinates (the proportions were estimated from Figure 20 of Egan, 1958), and the bottom panel shows the proportions on log-log coordinates.

First note that, irrespective of the form of the underlying distribution, d is larger after two repetitions of the list (the open symbols are farther from the diagonal). Thus, the plots show that repetition of the list increases memory strength, and this result appears to be robust with respect to the assumed form of the underlying distributions.

Next, consider the slopes of the ROC curves. The top panel of Figure 8 shows that, on logit coordinates, the curves have slopes less than unity, as is also the case for inverse normal coordinates (as shown by Egan, 1958); note that the curves are also approximately parallel. This has been interpreted as showing that the variance of the old word distribution is larger than that of the new word distribution and that repetition of the list increases memory strength but does not affect the relative variances. Ratcliff et al. (1992) discussed implications of this result for current memory theories.

The lower panel of the figure shows that, on log-log coordinates, the ROC curves have slopes either close to unity (for one repetition) or slightly greater than unity (for two repetitions). This can be interpreted as showing that the extreme value signal and noise distributions have equal variances and that repetition of the list increases memory strength, with the signal variance possibly becoming relatively smaller (another possibility is that the signal distribution is relatively less skewed after two repetitions; see gen-

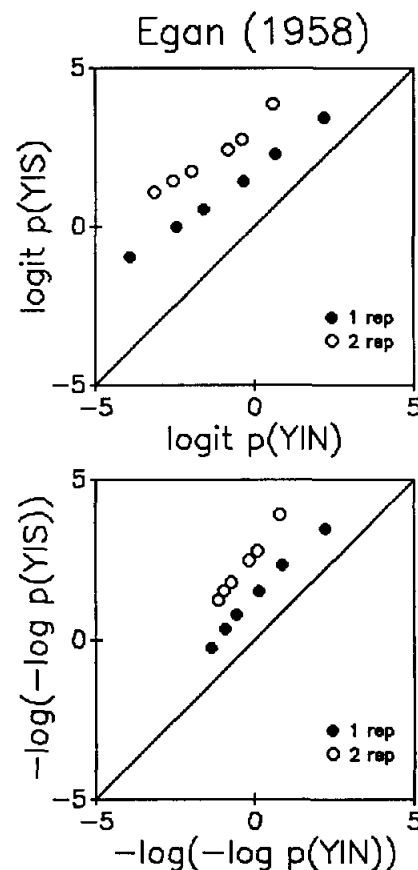


Figure 8. Logistic (upper panel) and extreme value (lower panel) receiver operating characteristic plots for a rating response word recognition experiment of Egan (1958, Figure 20). The filled circles are the average proportions for 8 participants who saw the study list once before being tested; the open circles are for a different group of 8 participants, who saw the study list twice before being tested; $p(Y|N)$ is the probability of reporting a signal (Y) given that noise (N) was presented (a false alarm); $p(Y|S)$ is the probability of reporting a signal given that a signal (S) was presented (a hit). rep = repetition.

eralized probability distributions on p. 198). Thus, conclusions about the relative variances can differ, depending on the assumed form of the underlying distributions. Researchers should keep this in mind when formulating or revising theories.

On Some Extended Signal Detection Models

An advantage of a model-based approach to SDT is that it offers the power and flexibility of a regression-like framework. Just as regression models have been extended in various ways, such as to allow for richer

covariance structures in time series and longitudinal models, GLMs have been similarly generalized, and so the SDT models discussed above can be extended. In this section I comment on some extended models and suggest avenues for future research.

Dynamic Signal Detection Models

All of the traditional signal detection models assume independent responses. However, the data are often repeated responses from participants, and so the responses might be correlated. There are several ways to approach possible dependencies among the responses; the choice depends in part on the goals of the research, the type of data available (e.g., time series or longitudinal data), and on whether the response correlation is of theoretical interest or is regarded as a nuisance. For example, in signal detection research, one often obtains many responses from each participant, so the data are time series, and conditional models might be useful for exploring the correlation structure. In other situations, as in applied research, only a small number of responses may be obtained from each participant, but there are possibly many participants, so the data are longitudinal, and mixed or marginal models might be useful.

Conditional models. In some situations, the correlation of responses is of theoretical interest, because it might provide information about underlying psychological processes (e.g., DeCarlo, 1994). Several researchers have suggested, for example, that the response criterion in signal detection systematically varies over trials (e.g., Treisman & Williams, 1984), being driven by the previous response (or feedback). This notion can be used to motivate a signal detection model that includes the previous response as a regressor. This type of model is known variously as a *conditional*, *transitional*, or *Markov* model, because the response is modeled conditional on values of the previous response. Cox and Snell (1989) discussed conditional models for logistic regression, and Fahrmeir and Tutz (1994) discussed them for GLMs. The models remain to be explored in signal detection research; they might also be useful for research on random response generation.

Mixed models. Mixed models (called *mixed* because they include both fixed and random effects) allow one, among other things, to specify a covariance structure for the repeated measures. Mixed models have been extended to GLMs, resulting in generalized linear mixed models (e.g., Breslow & Clayton, 1995; Littell, Milliken, Stroup, & Wolfinger, 1996), and

they can also be used for signal detection models. For example, a binary response signal detection model with a correlation structure for the responses, such as first-order autoregressive, can be fit by using the SAS GLIMMIX macro (SAS Institute, 1989). Other software for mixed or hierarchical GLMs also is available (Hedeker & Gibbons, 1996; Kreft, de Leeuw, & van der Leeden, 1994), and it can be used for SDT.

Mixed models also allow one to consider models with random coefficients, which are also known as *hierarchical* or *multilevel* models. Swets (1992, p. 530) gave an example in SDT where an approach via hierarchical models might be useful. Technicians made binary decisions as to the presence or absence of metal fatigue; the 254 technicians were from 17 Air Force bases. The hierarchical structure of the design (technicians nested within bases) can be recognized by a signal detection model where d (and possibly c) is a random coefficient across the bases (and variation within and across bases can be examined).

Marginal models. In some situations, interest centers on the signal detection parameters, usually d , and any correlation of the responses is regarded as a nuisance. A concern in this case is that ignoring the correlation can affect, for example, standard errors and confidence intervals for the parameter estimates. Marginal models, which separately model the (marginal) response probabilities and the association among the repeated responses, might be useful in this situation (for longitudinal data). For example, Liang and Zeger (1986) introduced *generalized estimating equations* (GEEs) to take the response correlation into account (Dunlop, 1994, provided an introduction). The approach yields consistent estimates of the (population average) parameters and their standard errors with only mild assumptions about the correlation structure. Marginal models have been discussed by, among others, Fahrmeir and Tutz (1994) and Diggle, Liang, and Zeger (1994); SAS and SPSS macros for GEEs also are available. Again, the utility of marginal models for signal detection research remains to be investigated.

In the next two sections I examine extensions of the basic signal detection model. The first involves introducing a parameter to allow the variances of the underlying distributions to differ; the second introduces a parameter to allow the shape of the underlying distributions to vary.

Unequal Variance SDT Models

A traditional approach to nonunit slope ROC curves has been to introduce a parameter to allow the

variances of the signal and noise distributions to differ (see Green & Swets, 1966). For example, for a logistic model, the scale parameter τ is replaced by τ_s for the signal distribution and by τ_n for the noise distribution, and the model is

$$\text{logit } p(Y \leq j|X) = \frac{c_j - d_n X}{\tau_s^X},$$

where d_n is the distance measure scaled with respect to τ_n , referred to as Δm in the normal theory version of SDT (Green & Swets, 1966), and it is assumed without loss of generality that $\psi_n = 0$ and $\tau_n = 1$. Replacing the logit link with a general link gives a flexible class of unequal variance signal detection models (see Tosteson & Begg, 1988). The models are not GLMs because they are nonlinear in the parameters, but they can be fit by alternating between two submodels, one for the numerator of the above, and one for the denominator (nonlinear procedures can also be used).

The linearized ROC curve for the unequal variance model is

$$\text{logit } p(Y > j|S) = d_s + \frac{\tau_n}{\tau_s} \text{logit } p(Y > j|N),$$

where the intercept d_s is the distance measure scaled with respect to τ_s , and the slope is $\tau_n/\tau_s = \sigma_n/\sigma_s$. Thus, from the perspective of the unequal variance model, ROC curves with slopes less than unity indicate that $\sigma_n < \sigma_s$.

Generalized Probability Distributions

Another approach to nonunit slope ROC curves is to allow the shape (e.g., skew or kurtosis) of the underlying distributions to vary. For example, one approach is to use the log of a random variable distributed as Burr Type XII (see Johnson, Kotz, & Balakrishnan, 1994), which includes the logistic and extreme value distributions as special cases. The log Burr has been used in bioassay (see Morgan, 1992) and in SDT (DeCarlo, 1997).

More specifically, a signal detection model based on the log Burr distribution is

$$p(Y \leq j|X) = 1 - (1 + \lambda \exp(c_j - dX))^{-1/\lambda},$$

where the term on the right is the CDF for the log Burr, and λ is a shape parameter ($\lambda > 1$ gives positive skew, and $\lambda < 1$ gives negative skew). The inverse of the above gives the model written as a GLM,

$$\log \frac{(1-p)^{-\lambda} - 1}{\lambda} = c_j - dX.$$

The above shows that the underlying probability distribution is generalized by including a parameter in the link function of the GLM. The logit link is obtained for $\lambda = 1$, and $\lambda = 0$ gives the complementary log-log link, so the logistic and extreme value SDT models are special cases. Note that the link is similar to the log of a Box-Cox (1964) power transformation, again keeping in mind that GLMs accomplish the transformation through the link function.

The log Burr SDT model can be fit by using software for GLMs that allow for a user-defined link function; maximum likelihood estimates of the parameters can be obtained by performing a search over λ . For example, the GENMOD procedure of SAS (SAS Institute, 1989) can be used (DeCarlo, 1997). The standard errors, however, are incorrect because λ is assumed to be known; Morgan (1992, Appendix C) showed how to obtain adjusted (asymptotic) standard errors.

In sum, the regressionlike approach offered by GLMs can be extended in a variety of ways, each of which offers interesting possibilities for SDT. The models also have implications for other research areas; they suggest, for example, similar generalizations for item response theory models (cf. Mellenbergh, 1994) and also suggest that it might be worthwhile to consider alternatives to the multivariate probit models used in structural equation modeling (e.g., Muthén, 1984).

Conclusions

GLMs offer a unified framework for signal detection models, from the simplest to the most complex. The approach allows researchers with a good working knowledge of regression analysis to examine a variety of signal detection models. One can consider, for example, models based on different underlying distributions, models that allow for response dependencies, and models with random coefficients, to name a few possibilities. Recent advances in statistics can also be used, such as, in addition to those noted above, software for exact logistic regression (Mehta & Patel, 1996), which might be useful for small sample sizes, and power analysis, bootstrapping, and jackknifing for GLMs, all of which can be brought to bear on signal detection research.

I hope that the present article will encourage researchers to use the above models in their research.

The result will be a deeper and more informative analysis of signal detection data.

References

- Agresti, A. (1989). Tutorial on modeling ordered categorical response data. *Psychological Bulletin*, 105, 290–301.
- Agresti, A. (1990). *Categorical data analysis*. New York: Wiley.
- Agresti, A. (1996). *An introduction to categorical data analysis*. New York: Wiley.
- Box, G. E. P., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society, Series B*, 26, 211–243.
- Breslow, N. E., & Clayton, D. G. (1995). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, 88, 9–25.
- Camilli, G. (1994). Origin of the scaling constant $d = 1.7$ in item response theory. *Journal of Educational and Behavioral Statistics*, 19, 293–295.
- Coombs, C. H., Dawes, R. M., & Tversky, A. (1970). *Mathematical psychology: An elementary introduction*. Englewood Cliffs, NJ: Prentice Hall.
- Cox, D. R., & Snell, E. J. (1989). *Analysis of binary data* (2nd ed.). New York: Chapman and Hall.
- DeCarlo, L. T. (1994). A dynamic theory of proportional judgment: Context and judgment of length, heaviness, and roughness. *Journal of Experimental Psychology: Human Perception and Performance*, 20, 372–381.
- DeCarlo, L. T. (July 1997). *Generalized signal detection models via generalized linear models*. Paper presented at the 10th European meeting of the Psychometric Society, Santiago de Compostela, Spain.
- Diggle, P. J., Liang, K., & Zeger, S. L. (1994). *Analysis of longitudinal data*. New York: Oxford University Press.
- Dobson, A. J. (1990). *An introduction to generalized linear models*. New York: Chapman and Hall.
- Dorfman, D. D., & Alf, E., Jr. (1968). Maximum likelihood estimation of parameters of signal detection theory—A direct solution. *Psychometrika*, 33, 117–124.
- Dunlop, D. D. (1994). Regression for longitudinal data: A bridge from least-squares regression. *The American Statistician*, 48, 299–303.
- Egan, J. P. (1958). *Recognition memory and the operating characteristic* (Technical Note No. AFCRC-TN-58-51, AO-152650). Bloomington: Indiana University Hearing and Communication Laboratory.
- Egan, J. P. (1975). *Signal detection theory and ROC analysis*. New York: Academic Press.
- Fahrmeir, L., & Tutz, G. (1994). *Multivariate statistical modelling based on generalized linear models*. New York: Springer-Verlag.
- Gescheider, G. A. (1997). *Psychophysics: The fundamentals* (3rd ed.). Hillsdale, NJ: Erlbaum.
- Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.
- Hasselblad, V., & Hedges, L. V. (1995). Meta-analysis of screening and diagnostic tests. *Psychological Bulletin*, 117, 167–178.
- Hauck, W. W., & Donner, A. (1977). Wald's test as applied to hypotheses in logit analysis. *Journal of the American Statistical Association*, 72, 851–853.
- Hedeker, D., & Gibbons, R. D. (1996). MIXOR: A computer program for mixed-effects ordinal regression analysis. *Computer Methods and Programs in Biomedicine*, 49, 157–176.
- Hilbe, J. M. (1994). Generalized linear models. *The American Statistician*, 48, 255–265.
- Hosmer, D. W., & Lemeshow, S. (1989). *Applied logistic regression*. New York: Wiley.
- Johnson, N. L., Kotz, S., & Balakrishnan, N. (1994). *Continuous univariate distributions* (2nd ed., Vol. 1). New York: Wiley.
- Johnson, N. L., Kotz, S., & Balakrishnan, N. (1995). *Continuous univariate distributions* (2nd ed., Vol. 2). New York: Wiley.
- Klein, J. P., & Moeschberger, M. L. (1997). *Survival analysis: Techniques for censored and truncated data*. New York: Springer-Verlag.
- Kleinbaum, D. G. (1994). *Logistic regression: A self-learning text*. New York: Springer-Verlag.
- Kreft, I. G. G., de Leeuw, J., & van der Leeden, R. (1994). Review of five multilevel analysis programs: BMDP-5V, GENMOD, HLM, ML3, and VARCL. *The American Statistician*, 48, 324–335.
- Lawless, J. F. (1982). *Statistical models and methods for lifetime data*. New York: Wiley.
- Liang, K., & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, 73, 13–22.
- Liao, T. F. (1994). *Interpreting probability models: Logit, probit, and other generalized linear models*. London: Sage.
- Littell, R. C., Milliken, G. A., Stroup, W. W., & Wolfinger, R. D. (1996). *SAS system for mixed models*. Cary, NC: SAS Institute, Inc.
- MacMillan, N. A., & Creelman, C. D. (1991). *Detection theory: A user's guide*. New York: Cambridge University Press.
- McCullagh, P. (1980). Regression models for ordinal data. *Journal of the Royal Statistical Society, Series B*, 42, 109–142.
- McCullagh, P., & Nelder, J. A. (1989). *Generalized linear models* (2nd ed.). New York: Chapman and Hall.

- McNicol, D. (1972). *A primer of signal detection theory*. London: Allen & Unwin.
- Mehta, C., & Patel, N. (1996). *LogXact for Windows*. Cambridge, MA: CYTEL Software Corporation.
- Mellenbergh, G. J. (1994). Generalized linear item response theory. *Psychological Bulletin*, 115, 300–307.
- Morgan, B. J. T. (1976). The uniform distribution in signal detection theory. *British Journal of Mathematical and Statistical Psychology*, 29, 81–88.
- Morgan, B. J. T. (1992). *Analysis of quantal response data*. New York: Chapman and Hall.
- Muthén, B. (1984). A general structural equation model with dichotomous, categorical, and continuous latent variable indicators. *Psychometrika*, 49, 115–132.
- Nelder, J. A., & Wedderburn, R. W. M. (1972). Generalized linear models. *Journal of the Royal Statistical Society, Series A*, 135, 370–384.
- Noreen, D. L. (1977, August). *Relations among some models of choice*. Paper presented at the 10th Annual Mathematical Society Meeting, University of California, Los Angeles.
- Ogilvie, J. C., & Creelman, C. D. (1968). Maximum-likelihood estimation of receiver operating characteristic curve parameters. *Journal of Mathematical Psychology*, 5, 377–391.
- Rao, C. R. (1948). Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation. *Proceedings of the Cambridge Philosophical Society*, 44, 50–57.
- Ratcliff, R., Sheu, C., & Gronlund, S. D. (1992). Testing global memory models using ROC curves. *Psychological Review*, 99, 518–535.
- SAS Institute, Inc. (1989). *SAS/STAT user's guide* (Version 6, 4th ed.). Cary, NC: Author.
- Singh, S. N., & Churchill, G. A., Jr. (1986). Using the theory of signal detection to improve ad recognition testing. *Journal of Marketing Research*, 23, 327–336.
- SPSS, Inc. (1994). *SPSS guide to advanced statistics*. Chicago: Author.
- Strauss, D. (1992). The many faces of logistic regression. *The American Statistician*, 46, 321–327.
- Swets, J. A. (December, 1973). The relative operating characteristic in psychology. *Science*, 182, 990–1000.
- 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. (1992). The science of choosing the right decision threshold in high-stakes diagnostics. *American Psychologist*, 47, 522–532.
- 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.
- Tanner, W. P., Jr., & Swets, J. A. (1954). A decision-making theory of visual detection. *Psychological Review*, 61, 401–409.
- Tosteson, A. N. A., & Begg, C. B. (1988). A general regression methodology for ROC curve estimation. *Medical Decision Making*, 8, 204–215.
- Treisman, M., & Williams, T. C. (1984). A theory of criterion setting with an application to sequential dependencies. *Psychological Review*, 91, 68–111.
- Wandell, B., & Luce, R. D. (1978). Pooling peripheral information: Averages versus extreme values. *Journal of Mathematical Psychology*, 17, 220–235.

Appendix A

Notes on SDT and GLMs

The Log Odds Ratio

The distance measure d for the logistic SDT model is also a log *odds ratio* (OR), an important measure of association widely used in biostatistics (see Agresti, 1990). The log OR is obtained by taking the difference between the two logits shown in Equation 1, from which it follows that the log OR = d . Thus, the estimate of $d_1 = 1.60$ for the not-hypnotized participant, discussed below, shows that the odds of saying "yes" to an old word were $\exp(1.60) = 4.95$ times larger than the odds of saying "yes" to a new word.

A Choice Theory Parameterization

An alternative parameterization of the model uses *effect* coding, in which a signal is coded as 1 and noise is coded as -1. This gives a choice theory version of the model (e.g., Noreen, 1977). In particular, Equation 2 can be written as

$$\text{logit } p(Y = 1|X) = \frac{c - (\psi_s + \psi_n)/2}{\tau} + \frac{\psi_s - \psi_n}{2\tau} X.$$

Note that, when $X = 1$, the above gives the log odds of a hit, as shown in Equation 1, and when $X = -1$ it gives the log odds of a false alarm. The intersection point of the two symmetrical distributions is at $(\psi_s + \psi_n)/2$; setting 0 at the intersection point and $\tau = 1$ gives

$$\text{logit } p(Y = 1|X) = -c' + (d/2)X.$$

The coefficient of X in this case is $d/2$, which is simply the logarithm of the choice theory parameter α , and the intercept times -1 is the distance of the criterion from the intersection point, say c' , and is the logarithm of the choice theory bias parameter b ; note that $c' = c - (1/2)d$. For example, for the hypnotized participant discussed below, the obtained estimates are $d/2 = \log(\alpha) = 0.86$ and $c' = \log(b) = -1.23$, which match those shown in the lower panel of Figure 2.3 of MacMillan and Creelman (1991, p. 36).

Classification and Identification

Equation 4 can be used for binary or rating response experiments in which several stimuli are presented, as in classification or identification experiments, by using dummy variables for the multiple

stimuli. Appendix B presents a sample SAS program (SAS Institute, 1989) for a rating response intensity identification experiment presented in MacMillan and Creelman (1991) in Table 9.3 on p. 222. In this case, using the normit link, -1 times the coefficients of X (stim2, stim3, and stim4) gives estimates of d_{12} , d_{13} , and d_{14} (Stimulus 1 is used as the reference), which are 1.19, 1.57, and 2.38, respectively. It follows that estimates of d_{12} , d_{23} , and d_{34} are 1.19, 0.38, and 0.81, which match MacMillan and Creelman's (1991) reported estimates (from fits by eye!) of 1.2, 0.4, and 0.8 (p. 222).

A Binary Response Example

The example is from Chapter 2 of MacMillan and Creelman (Example 2a, p. 31). The data consist of two 2×2 tables from a (hypothetical) face recognition experiment with two conditions: a not-hypnotized participant and a hypnotized participant (the tables are treated as independent). The two participants are shown faces, and the task is to determine during a test whether the face has been shown before in the experiment. The parameter d in this case is a measure of recognition memory strength. Appendix B presents the data and sample SAS (SAS Institute, 1989) and SPSS (SPSS, 1994) programs.

The basic approach is to fit Equation 2 to the data (hits and false alarms) and to use dummy variables to allow the signal detection parameters to differ across the participants. As for ANOVA and regression models, the model can be parameterized in more than one way. One approach is to use the not-hypnotized participant as a reference and to use a dummy variable to allow the parameters to differ for the hypnotized participant. More specifically, a variable (X_1) indicating the condition (0 = not hypnotized, 1 = hypnotized) allows c to differ for the hypnosis condition, and a product term (condition \times signal, $X_1 \times X_2$) allows d to differ (see the program in Appendix B), and the model is

$$\text{logit } p(Y > j|X) = -c_1 - (c_2 - c_1)X_1 + d_1X_2 + (d_2 - d_1)X_1 \times X_2.$$

Using this approach, the coefficient of the signal regressor X_2 is an estimate of d_1 for the reference participant, which is 1.60, and -1 times the intercept

is an estimate of c_1 , which is 0.80. The coefficient of the product term $X1 \times X2$ is an estimate of $d_2 - d_1$, which is 0.13, and the coefficient of the condition variable $X1$ is an estimate of $-(c_2 - c_1)$, which is 1.16. Thus, d_2 for the hypnotized participant is $1.6 + 0.13 = 1.73$, and c_2 is $0.80 + (-1.16) = -0.36$.

The Wald statistic for each regression coefficient tests the null hypothesis that the coefficient is 0. It is computed as the square of the ratio of the estimated coefficient to its standard error and is asymptotically distributed under the null hypothesis as a chi-square with 1 *df*; it is analogous to the *t* statistic used to test coefficients in ordinary regression (Hauck & Donner, 1977, noted limitations of the Wald test). In this example, the test for the signal \times condition product variable tests the null hypothesis of equal detection parameters across the two conditions, $H_0: d_2 - d_1 = 0$, which is not rejected ($df = 1$, $p = .7949$).

A likelihood ratio (LR) test of the null hypothesis $H_0: d_2 - d_1 = 0$ can be performed by fitting two models: a full model that includes the condition, signal, and product term as regressors, and a reduced model that does not include the product term, so that $d_2 = d_1$ (note that c is allowed to vary across the two conditions). The LR test is based on the difference between -2 times the maximized log likelihoods ($-2 \log L$) obtained for a fit of the reduced and full models, respectively, with degrees of freedom equal to the difference in the number of parameters across the two models. For this example, the LR statistic is 0.068 with 1 *df* and $p = .794$, so the null hypothesis $H_0: d_2 - d_1 = 0$ is not rejected. A further reduced model with both $d_1 = d_2$ and $c_1 = c_2$ gives an LR statistic of 28.455 with 1 *df* and $p = .000$. Thus, recognition memory does not differ across the two conditions, but the response criterion does.

One can also code dummy variables so that direct estimates of the signal detection parameters and their standard errors are obtained; this is shown in Appendix B with the GENMOD procedure of SAS. The model is fit without an intercept, and the program shows how to use contrast statements to test hypotheses about the parameters. Yet another option is to use effect coding, which gives a choice theory parameterization of the model, as shown by the program in Appendix B.

A Rating Response Example

Ogilvie and Creelman (1968) provided the raw data for a single participant in a rating response experiment

on two-point touch sensitivity (the two-point separation was 0.5 in., and testing was done on the forearm). Appendix B provides the data and a SAS Program.

The first step is to examine the ROC plot, from which it appears that the data lie on an ROC curve with a slope of unity. The maximum likelihood estimate of d is 1.62 with a standard error of 0.17. An LR test of the null hypothesis of proportional odds, $H_0: d_1 = d_2 = d_3 = d_4 = d_5$, gives a likelihood ratio statistic of 2.66 with 4 *df* and $p = .617$, which is not significant. SAS also provides a *score* test (Rao, 1948), known in econometrics as the *Lagrange multiplier* test, of the proportional odds assumption. The score test is asymptotically equivalent to the LR test under the null hypothesis but requires only a fit of the restricted model, whereas the LR test requires a fit of both the restricted and unrestricted models. Both the LR and score tests show that the proportional odds assumption is not rejected, which suggests that the data lie on an ROC curve with a slope of unity, as expected in light of the ROC plot.

Note that rejection of the null hypothesis of proportional odds could occur for several reasons, because the alternative hypothesis is simply that the odds are not proportional. One possibility, for example, is that the data lie on the same ROC curve, but the curve has a slope other than unity; another possibility is that the data do not lie on the same ROC curve. Visual inspection of the empirical ROC curve is useful for assessing these alternatives. In addition, a test of a unit slope ROC curve against a nonunit slope alternative can be performed by fitting an unequal variance signal detection model and comparing the $-2 \log$ likelihood to that of the equal variance model.

The programs also provide goodness-of-fit statistics: an LR statistic and a Pearson chi-square statistic (Ogilvie & Creelman, 1968, reported the Pearson chi-square). Goodness-of-fit statistics indicate how close the predicted values are to the observed values; they provide a rough assessment of how well the model fits the data. It should be kept in mind that goodness-of-fit statistics are affected by sample size and that components of fit, such as residuals and other diagnostics, should also be examined.

The LR goodness-of-fit statistic, as for the LR test discussed above, involves a comparison of two models: a reduced model and a saturated model,

$$LR X^2 = (-2\log L_R) - (-2\log L_S),$$

where L_S is the likelihood for the saturated model.

The saturated model has as many parameters as observations, so it fits perfectly and exactly reproduces the observations (it provides a baseline for comparison). A comparison of the reduced model to the saturated model assesses how close the predicted values are to the observed values, which reflects the goodness of fit.

For the rating response example, the goodness-of-fit statistics are close to their degrees of freedom, indicating an acceptable fit. Note that the test of proportional odds is equivalent in this example (but not in general) to the tests of goodness of fit, because both tests assess the fit of the model with constant d relative to a saturated model.

Appendix B

Programs for Signal Detection Examples

SAS Programs

```
options ls=80; options formdlim='—';
title 'Example 2a from MacMillan & Creelman, 1991, p. 31';
*(Note—the events/trials syntax is used for the response.
  See the SAS/STATS user's guide for details);
data temp;
  input cond $ hypno signal pyes trials;
  sighypno = signal*hypno;
  cards;
    normal 0 0 31 100
    normal 0 1 69 100
    hypnot 1 0 59 100
    hypnot 1 1 89 100
  ;
title2 'Fit of the saturated model with normal as reference';
proc logistic data=temp;
  model pyes/trials=hypno signal sighypno;
run;
title2 'Reduced model with d equal across the two conditions';
proc logistic data=temp;
  model pyes/trials=hypno signal / scale=none;
run;
title2 'An example of PROC PROBIT';
proc probit data=temp;
  model pyes/trials=hypno signal/d=logistic lackfit;
run;
title2 'An example of PROC GENMOD and contrast statements';
*(Note—the model is fit without an intercept, and the
  coefficients give direct estimates of  $d_j$  and  $-c_j$ );
proc genmod data=temp order=data;
  class cond;
  model pyes/trials=cond cond*signal/noint dist=bin link=logit;
  contrast 'd1=d2' cond*signal 1 -1;
run;
title2 'A fit of the choice theory version of the model';
* the results match those shown by MacMillan & Creelman, p.36;
data choice;
  input cnorm chypno dnorm dhypno pyes trials;
  cards;
```

```

1 0 -1 0 31 100
1 0 1 0 69 100
0 1 0 -1 59 100
0 1 0 1 89 100
;
proc logistic data = choice;
  model pyes/trials = cnorm chypno dnorm dhypno/noit;
run;
*****
title "Rating exp. - Ogilvie & Creelman, J. Math. Psy, 1968";
* responses are coded 1 = sure noise to 6 = sure signal;
data rate;
  input stim resp freq @@;
  cards;
    0 6 15 0 5 17 0 4 40 0 3 83 0 2 29 0 1 66
    1 6 68 1 5 37 1 4 68 1 3 46 1 2 10 1 1 21
  ;
title2 "Fit of the logistic model using PROC LOGISTIC";
proc logistic data = rate;
  weight freq;
  model resp = stim / aggregate scale = none;
run;
*****
title "Intensity identification - MacMillan & Creelman, p.222";
*(Note—the reference distribution corresponds to stimulus 1);
data ident;
  input stim2 stim3 stim4 resp freq @@;
  cards
    0 0 0 1 39 0 0 0 2 7 0 0 0 3 3 0 0 0 4 1
    1 0 0 1 17 1 0 0 2 12 1 0 0 3 10 1 0 0 4 11
    0 1 0 1 11 0 1 0 2 10 0 1 0 3 12 0 1 0 4 17
    0 0 1 1 3 0 0 1 2 5 0 0 1 3 9 0 0 1 4 33
  ;
proc logistic data = ident;
  weight freq;
  model resp = stim2 stim3 stim4/link = normit;
run;
*****
title "Obs 1, rating response, Swets, Tanner, & Birdsall 1961";
*(Note—responses are coded 1 = sure noise to 6 = sure signal);
data temp;
  input signal resp freq @@;
  cards;
    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
  ;
title2 "Fit of the logistic model";
proc logistic data = temp;
  weight freq;
  model resp = signal/link = logit aggregate scale = none;
run;
title2 "Fit of the extreme value model";
proc logistic data = temp;
  weight freq;
  model resp = signal/link = cloglog aggregate scale = none;
run;

```

An SPSS Program

```

set width=80 length=none.
title "Example 2a from MacMillan & Creelman, 1991, p.31".
*(Note—responses are coded yes = 1 and no = 0).
data list list /hypno signal yes count *.
begin data
0 0 0 69
0 0 1 31
0 1 0 31
0 1 1 69
1 0 0 41
1 0 1 59
1 1 0 11
1 1 1 89
end data.
compute sighypno = signal*hypno.
weight by count.
logistic regression yes with hypno signal sighypno
/criteria lcon(0).
* Next is the restricted model without the interaction term.
* The -2 log L can be used to test for constant dj.
logistic regression yes with hypno signal
/criteria lcon(0).

```

Received June 13, 1997

Accepted October 3, 1997 ■