Chapter 12: Judgment and Choice

Prerequisites: Chapter 5, Sections 3.9, 3.10, 6.8

12.1 *Historical Antecedents*

In the 19th century Gustav Fechner attempted to understand how it is that humans perceive their world.   The simplest place to start was by asking how it is that we perceive basic physical quantities such as the heaviness of a block of wood, the brightness of a light, or the loudness of a tone.   He thought that there were three important elements behind the sequence by which the process operates:

(1) The external physical environment, which we will denote n
(2) Brain activity, which we will denote m, and
(3) Conscious perception, which we will denote s.

Fechner believed that the relationship between (2) and (3) was inaccessible to science, and that anyway, they were just two different ways of looking at the same phenomenon. On the other hand, the relationship between (1) and (2) was part of physics, or perhaps physiology.   Here, he concluded that there was some sort of one-to-one correspondence.   He decided that he would investigate the relationship between (1) and (3), and, some would argue, by doing so created the science that we call psychology.   He was concerned therefore with the way that simple physical stimuli come to be perceived.   He proposed the following law, now known as *Fechner's Law*:

$$s = c \ln [n / n_0] \tag{12.1}$$

where s has been previously defined as the conscious perception of the loudness, brightness, or heaviness in question; and n the actual physical value of the stimulus.   The constant c summarizes the sensitivity of the sense in question, while n_0 is the *absolute threshold*.   The absolute threshold is the lowest limit of perception.   For example, if we are talking about sounds, n_0 would be the softest sound detectable.   The fact that Fechner used a log function is particularly meaningful.   We can relate this to a variety of concepts, such as the economic notion of diminishing returns.   The function predicts that proportional changes are equally important.   In other words, if I am holding a one ounce block and I add 1/10th of an ounce of additional weight, this creates the same amount of perceived change as if I had a 1 pound block and I add 1/10th of a pound.   This notion was later empirically verified by Weber who discovered that the size of a just noticeable difference was proportion to n,

$$\Delta n = kn \tag{12.2}$$

where k quantifies the sensitivity of the sense for the observer.

We now continue this historical review with the notion of absolute detection.   We will say that the physical stimulus is measured in units of n, for example seconds, kilograms, centimeters, foot-candles, and so forth.   In the 19th century it was imagined that there was a threshold, above which perception of the stimulus began, and below which there was no perception.   Assuming we are dealing with brightness, it was assumed that as n increased, the conscious perception of the light popped suddenly into existence:
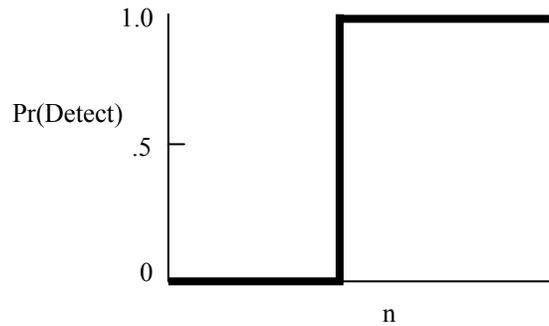
The position where this occurred was called the absolute threshold. A related experiment might have subjects compare two lights, and to make a judgment as to which was brighter. Then the question became one of *difference thresholds*, that is, a point above which the comparison light would be perceived of as identical and below which it would be perceived as dimmer, and another point, above which the comparison would be seen as brighter. The situation is pictured below.
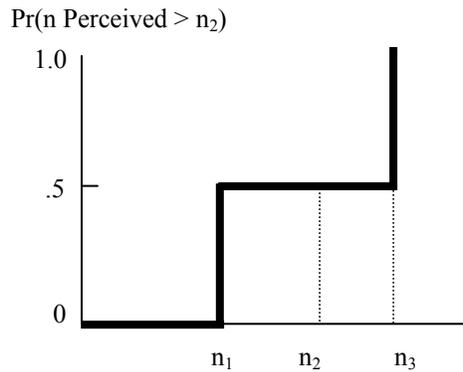
We would say that the upward JND (Just Noticeable Difference) would be the interval $n_3 - n_2$ and the downward JND would be $n_2 - n_1$. Things did not turn out like the graphs pictured above, however. In fact, empirical data for the probability of detection revealed a much smoother function. An idealized example is given below:
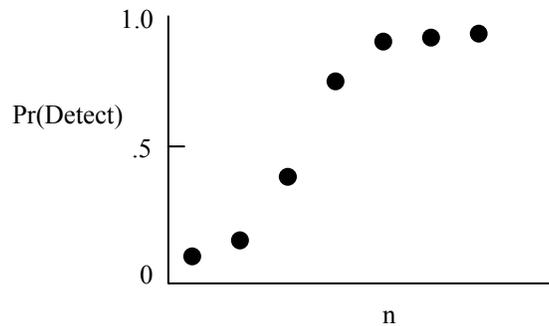
How can we account for this?

12.2 *A Simple Model for Detecting Something*

Here we propose a simple model that says the psychological effect of a stimulus i is

$$s_i = \bar{s}_i + e_i \tag{12.3}$$

where $\bar{s}_i$ is the impact on the sense organ of the observer and e_i is random noise, perhaps added by the nervous system, the senses or by distraction. Let us assume further that, as in Section 4.2,

$$e_i \sim N(0, \sigma^2) \quad \text{so that}$$

$$s_i \sim N(\bar{s}_i, \sigma^2) . \tag{12.4}$$

Now, assume that there actually is a fixed threshold so that the subject detects the stimulus if $s_i \geq s_0$, i. e. the threshold is located at s_0. More formally we can write that

$$\text{Pr[Detect stimulus i]} = \hat{p}_i = \text{Pr}[s_i \geq s_0] . \tag{12.5}$$

At this point we need to establish a zero point for the psychological continuum, s, that we have created. It would be convenient if we set $s_0 = 0$. This psychological continuum is of course no more than an interval scale, and so its origin is arbitrary. We might as well place the zero point at a convenient place. In that case, we have

$$\hat{p}_i = \frac{1}{\sqrt{2\pi}\sigma} \int\limits_0^{+\infty} \exp\left[-(s_i - \bar{s}_i)^2 / 2\sigma^2\right] ds_i . \tag{12.6}$$

Now we define $z = \dfrac{s_i - \bar{s}_i}{\sigma}$. In that case $dz/ds_i = 1/\sigma$ or $dz = ds_i/\sigma$. This will allow us to change the variable of integration, or in simple terms, switch everything to a standardized, z-score notation. This is shown below:

$$\hat{p}_i = \frac{1}{\sqrt{2\pi}\sigma} \int\limits_0^{+\infty} \exp\left[-(s_i - \bar{s}_i)^2 / 2\sigma^2\right] ds_i$$

$$= \frac{1}{\sqrt{2\pi}} \int\limits_{\frac{0-\bar{s}_i}{\sigma}}^{+\infty} \exp\left[\frac{-z^2}{2}\right] dz.$$

In words, as we go from the first line above to the last, we change from an "s-score" to a standardized z-score. In the first and second lines the integration begins at 0, but in the third line we have standardized so that we have subtracted the mean (from 0) and divided by σ. One last little change and we will have a very compact way to represent this probability. Since the normal distribution is symmetric, the area from $+z$ to $+\infty$ is identical to the area between $-\infty$ and $-z$. In

terms of the equation above, the area between $\dfrac{0 - \bar{s}_i}{\sigma}$ and $+\infty$ is then the same as that between $-\infty$

and $\dfrac{\bar{s}_i}{\sigma}$. We can therefore rewrite our detection probability as

$$\hat{p}_i = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{\bar{s}_i}{\sigma}} \exp\left[\frac{-z^2}{2}\right] dz$$

(12.7)

$$= \Phi[\bar{s}_i / \sigma],$$

where $\Phi(\cdot)$ is the standard normal distribution function [see Equation (4.14)]. A graphical representation of all of this appears below.
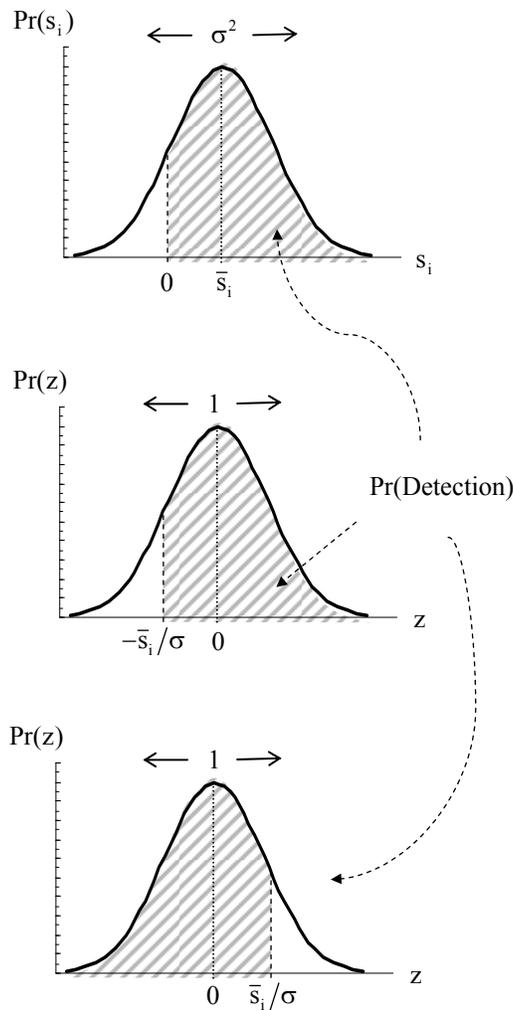
We can now summarize two points; one general and one particular to the detection problem at hand. In general, we might remember that for any random variable, call it a, for which $a \sim N$ [E(a), V(a)], then

$$\Pr [a \geq 0] = \Phi [E(a) / \sqrt{V(a)}] . \qquad (12.8)$$

In this particular case, s_i is playing the role of a, with $\bar{s}_i$ being E(a) and σ^2 being the V(a). And why do detection data not look like a step function? According to this model, they should look like a normal ogive. As the physical stimulus is varied, lets say by making it brighter and therefore easier to detect, $\bar{s}_i$ becomes larger and more and more of the distribution of s_i ends up being to the right of the threshold. This is illustrated in the figure below; with the shaded area representing the probability of detection of a light at three different intensities: dim, medium and bright.

$\bar{s}_1$ dim

$\bar{s}_2$ medium

$\bar{s}_3$ bright
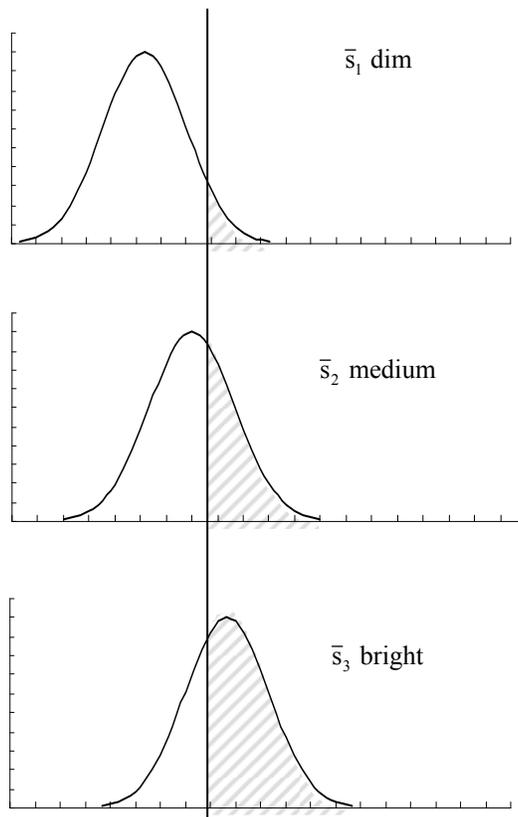
12.3 *Thurstone's Law of Comparative Judgment*

In the previous section we have discussed how people can detect something such as a dim light in a darkened room, a slight noise in an otherwise silent studio, or a small amount of a particular smell. That experimental situation is called absolute judgment, and we modeled it by positing the existence of a fixed threshold plus normal random noise. Now let's contemplate how people compare two objects, for example, which of two wooden blocks are heavier, a procedure known as comparative judgment. In 1927 L. L. Thurstone published a paper in which he specified a model

for the comparative process, generalizing the work that had gone on before by extending his analysis to stimuli that did not have a specific physical referent. His chosen example was "the excellence of handwriting specimens." This sort of example must stand alone, in the sense that we cannot rely on some sort of physical measure to help us quantify excellence. To Thurstone, that did not really matter. He simply proposed that for any property for which we can compare two objects, there is some psychological continuum. And the process by which we react differently to the several comparison objects is just called the "discriminal process."

We should not let this slightly anachronistic language throw us off. Thurstone's contribution was fundamental and highly applicable to 21st century marketing. Suppose I ask you to compare two soft drinks and to tell me which one you prefer. This is the situation that Thurstone addressed. We can use his method to create interval scale values for each of the compared brands, even though we are only collecting ordinal data: which of the two brands is preferred by each subject. This is the essence of psychological scaling – use the weakest possible assumptions (i. e. people can make ordinal judgments of their preferences) and still end up with interval level parameters. In the case of preference judgments, these parameters are usually called *utilities*, based on the economic theory of rational man.

To create an interval scale, Thurstone borrowed a data collection technique called *paired comparisons*. In paired comparisons, a subject makes a judgment on each unique pair of brands. For example, with four brands; A, B, C and D the subject compares A and B, AC, AD, BC, BD and CD. In general there are $q = \frac{t(t-1)}{2}$ unique pairs among t brands. An additional point should be added here. For one, it turns out that just looking at pairs is not the most efficient way to scale the t brands. Despite this, the mathematics behind Thurstone's Law is very instructive.

Lets look at a miniature example of paired comparison data. Consider the table below where a typical entry represents the probability that the row brand is chosen over the column brand.

	A	B	C
A	-	.6	.7
B	.4	-	.2
C	.3	.8	-

Each table entry gives the Pr[Row brand is chosen over the Column brand]. Such a table is sometimes called antisymmetric, as element i, j and element j, i must sum to 1.0. As such, we can use the q non-redundant pairs that appear in the lower triangular portion of the table as input to the model

Another point is that there are two different ways to collect data. If the brands are relatively confusable, you can collect data using a single subject. Otherwise, the proportions that appear in the table are aggregated over a sample of individuals.

As before, we will assume that the process of judgment of a particular brand, such as brand i, leads to an output, call it s_i. We further hypothesize that for brand i
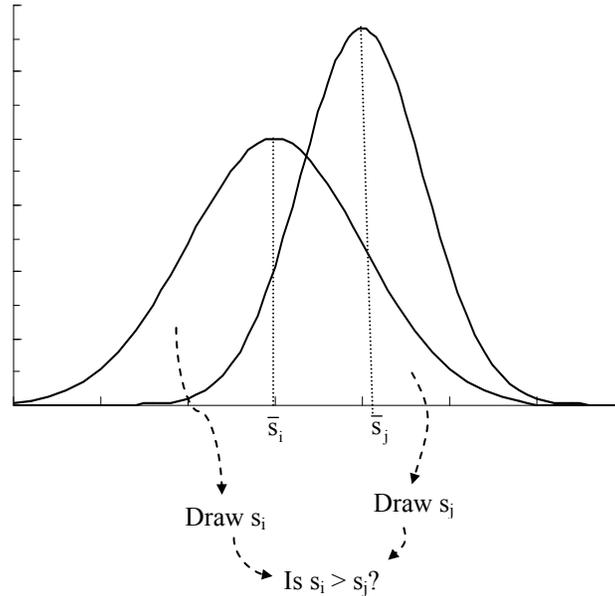
$$s_i = \bar{s}_i + e_i . \tag{12.9}$$

Similarly to what we did before in Equation (12.4), we further assume that

$$e_i \sim N(0, \sigma_i^2) \text{ with} \tag{12.10}$$

$$Cov(e_i, e_j) = \sigma_{ij} = r\sigma_i\sigma_j. \qquad (12.11)$$

We hypothesize that brand i is chosen over brand j whenever $s_i > s_j$. This situation, which as shown below, bears a certain resemblance to a two sample *t*-test:

Now we can say that the probability that brand i is chosen over brand j

$$p_{ij} = Pr(s_i > s_j) = Pr(s_i - s_j > 0).$$

So how will we derive that probability? Turning back a bit in this chapter, recall Equation (12.8) which gave us an expression for the $Pr(a > 0)$, namely $\Phi[E(a) / \sqrt{V(a)}]$, assuming that a is a normal variate. In the current case, the role of a is being played by $s_i - s_j$ and so we need to figure out $E(s_i - s_j)$ and $V(s_i - s_j)$. The expectation is simple.

$$E(s_i - s_j) = E\left[(\bar{s}_i + e_i) + (\bar{s}_j - e_j)\right]$$

$$= \bar{s}_i - \bar{s}_j,$$

since by our previous assumption the $E(e_i) = E(e_j) = 0$, and according to Theorem (4.4), the expectation of sum is the sum of the expectations. As far as the variance goes, we can use Theorem (4.9) to yield

$$V(s_i - s_j) = \begin{bmatrix} 1 & -1 \end{bmatrix} \begin{bmatrix} \sigma_i^2 & \sigma_{ij} \\ \sigma_{ij} & \sigma_j^2 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

$$= \sigma_i^2 + \sigma_j^2 - 2\sigma_{ij}$$

$$= \sigma_i^2 + \sigma_j^2 - 2r\sigma_i\sigma_j \, .$$

At this point we have all of the pieces that we need to figure out the probability that one brand is chosen over the other. It is

$$\hat{p}_{ij} = \Pr(s_i > s_j) = \Phi\left[(\bar{s}_i - \bar{s}_j) \Big/ \sqrt{\sigma_i^2 + \sigma_j^2 - 2r\sigma_i\sigma_j} \; \right]. \tag{12.12}$$

Thurstone imagined a variety of cases for this derivation. In Case I, one subject provides all of the data as we have mentioned before. In Case II, each subject judges each pair once and the probabilities are built up across a sample of different responses. In Case III, r is assumed to be 0 (or 1, it doesn't matter), and in Cases IV and V all of the variances are equal – exactly in Case V and approximately in Case IV.

12.4 *Estimation of the Parameters in Thurstone's Case III: Least Squares and ML*

We will continue assuming Case III, meaning that each brand can have a different variance, but the correlations or the covariances of the brands are identical. By convention we tie down the metric of the discriminal dimension, s, by setting $\bar{s}_1 = 0$ and $\sigma_1^2 = 1$. We will now look at four methods to estimate the $2(t-1)$ unknown parameters in the model, namely, the values $\bar{s}_2, \bar{s}_3, \cdots, \bar{s}_t, \sigma_2^2, \sigma_3^2, \cdots, \sigma_t^2$. These methods are unweighted nonlinear least squares, weighted nonlinear least squares, modified minimum χ^2 and maximum likelihood. Unweighted nonlinear least squares begins with the observation that with the model,

$$\Pr(s_i > s_j) = \Phi\left[(\bar{s}_i - \bar{s}_j) \Big/ \sqrt{\sigma_i^2 + \sigma_j^2} \; \right],$$

we can use the inverse normal distribution function, $\Phi^{-1}(\cdot)$ on both sides. To understand what Φ^{-1} does, lets remember what the Φ function does – remember that Φ is the standard normal distribution function. For example, $\Phi(1.96) = .975$, and $\Phi(0) = .5$. If Φ takes a z score and gives you the probability of observing that score or less, Φ^{-1} takes a probability and gives you a z score. So $\Phi^{-1}(.975) = 1.96$, for example. What this means is that if we transform our choice probabilities into z scores, we can fit them with a model that looks like

$$\Phi^{-1}[\Pr(s_i > s_j)] = \Phi^{-1}\left\{ \Phi\left[(\bar{s}_i - \bar{s}_j) \Big/ \sqrt{\sigma_i^2 + \sigma_j^2} \; \right] \right\}$$

$$\hat{z}_{ij} = (\bar{s}_i - \bar{s}_j) \Big/ \sqrt{\sigma_i^2 + \sigma_j^2} \, , \tag{12.13}$$

where $\hat{z}_{ij}$ is the predicted z score that corresponds to the choice that brand i is chosen over brand j. Now of course we have a string of such z scores, one for each of the q unique pairs,

$$\hat{z}_{12} = (\bar{s}_1 - \bar{s}_2)\Big/\sqrt{\sigma_1^2 + \sigma_2^2}$$

$$\hat{z}_{13} = (\bar{s}_1 - \bar{s}_3)\Big/\sqrt{\sigma_1^2 + \sigma_3^2}$$

$$\cdots = \cdots$$

$$\hat{z}_{(t-1)t} = (\bar{s}_{t-1} - \bar{s}_t)\Big/\sqrt{\sigma_{t-1}^2 + \sigma_t^2} \ .$$

In unweighted nonlinear least squares we will have as a goal the minimization of the following objective function –

$$f = \sum_{i=1}^{t-1} \sum_{j=i+1}^{t} (z_{ij} - \hat{z}_{ij})^2 \tag{12.14}$$

where the summation is over all q unique pairs of brands. This technique is called unweighted because it does not make any special assumptions about the errors of prediction, in particular, assuming that they are equal or homogeneous. In general, this assumption is not tenable when we are dealing with probabilities, but this method is quick and dirty and works rather well. We can use nonlinear optimization (see Section 3.9) to pick the various $\bar{s}_i$ and σ_i^2 values which are unknown a priori and must be estimated from the sample. We do this by picking starting values for each of the unknowns and then evaluating the vector of the derivative of the objective function with respect to each of those unknowns. We want to set this derivative vector to the null vector as below,

$$\begin{bmatrix} \partial f/\partial \bar{s}_1 \\ \partial f/\partial \bar{s}_2 \\ \cdots \\ \partial f/\partial \bar{s}_t \\ \hline \partial f/\partial \sigma_1^2 \\ \partial f/\partial \sigma_2^2 \\ \cdots \\ \partial f/\partial \sigma_t^2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \cdots \\ 0 \\ \hline 0 \\ 0 \\ \cdots \\ 0 \end{bmatrix}, \tag{12.15}$$

but we must do this iteratively, beginning with starting values and using these to evaluate the derivative. The derivative, or the slope, lets us know which way is "down", and we step off in that direction a given distance to come up with new, improved estimates. This process is repeated until the derivative is zero, meaning that we are at the bottom of the objective function, f.

The next approach also relies on nonlinear optimization and is called *weighted nonlinear least squares*, or in this case, it is also known as *Minimum Pearson χ^2*, since we will be minimizing the

classic Pearson Chi Square formula.   We will not be transforming the data using $\Phi^{-1}(\cdot)$.   Instead, we will leave everything as is, using the model formula

$$\hat{p}_{ij} = \Phi\left[(\bar{s}_i - \bar{s}_j)\Big/ \sqrt{\sigma_i^2 + \sigma_j^2} \right].$$

Our goal is to pick the $\bar{s}_i$ and the σ_i^2 so as to minimize

$$\hat{\chi}^2 = \sum_i^t \sum_{j \neq i}^t \frac{(np_{ij} - n\hat{p}_{ij})^2}{n\hat{p}_{ij}} \tag{12.16}$$

which the reader should recognize as the formula for the Pearson Chi Square with $n\hat{p}_{ij}$ a different way of writing the expected frequency for cell i, j. Note that in the above formula, the summation is over all off-diagonal cells and that $p_{ji} = 1 - p_{ij}$ and of course $\hat{p}_{ji} = 1 - \hat{p}_{ij}$. As an alternative, we can utilize matrix notation to write the objective function. This will make clear the fact that minimum Pearson Chi Square is a GLS procedure as discussed in Section 6.8, although in the current case our model is nonlinear. Now define

$$\mathbf{p}' = \begin{bmatrix} p_{12} & p_{13} & \cdots & p_{(t-1)t} \end{bmatrix}$$

and

$$\hat{\mathbf{p}}' = \begin{bmatrix} \hat{p}_{12} & \hat{p}_{13} & \cdots & \hat{p}_{(t-1)t} \end{bmatrix}.$$

Note also that for each element in $\mathbf{p}$,

$$V(p_{ij}) = V[p_{ij} - \hat{p}_{ij}] = \frac{\hat{p}_{ij}(1 - \hat{p}_{ij})}{n} \tag{12.17}$$

where n is the number of observations upon which the value p_{ij} is based. Using this information, we can create a diagonal matrix $\mathbf{V}$, placing each of the terms $\hat{p}_{ij}(1 - \hat{p}_{ij})\big/n$ on the diagonal of $\mathbf{V}$ in the same order that we placed the pair choice probabilities in $\mathbf{p}$ and $\hat{\mathbf{p}}$. In that case we can say

$$V(\mathbf{p}) = \mathbf{V}. \tag{12.18}$$

Now we will minimize

$$\hat{\chi}^2 = (\mathbf{p} - \hat{\mathbf{p}})' \mathbf{V}^{-1} (\mathbf{p} - \hat{\mathbf{p}}) \tag{12.19}$$

which is equivalent to the previous equation for Chi Square, and which is a special case of Equation (6.23). This technique is called weighted nonlinear least squares so as to distinguish it from ordinary, or unweighted, least squares. Also, remember that the elements in $\hat{\mathbf{p}}$, that is, the predicted pair choice probabilities, are nonlinear functions of the unknowns, the $\bar{s}_i$ and σ_i^2. For this reason we would use the nonlinear optimization methods of Section 3.9 here as well.

A third method we have of estimating the unknown parameters in the Thurstone model is called *Modified Minimum* χ^2 or sometimes *Logit* χ^2. In this case the objective function differs only slightly from the previous case, substituting the observed data for the expectation or prediction in the denominator:

$$\hat{\chi}^2 = \sum_i^t \sum_{j \neq i}^t \frac{(np_{ij} - n\hat{p}_{ij})^2}{np_{ij}} \ .$$

(12.20)

This tends to simplify the derivatives and the calculations somewhat, but perhaps is not as necessary as it once was when computer time was more expensive than it is today.

Before we turn to Maximum Likelihood Estimation, it could be noted here that we might also use a *Generalized Nonlinear Least Squares* approach that takes into account the covariances between different pairs (Christoffersson 1975, p. 29)

$$\text{Cov}(p_{ij}, p_{kl}) = \frac{p_{ijkl} - p_{ij} p_{kl}}{n}$$

(12.21)

where p_{ijkl} is the probability that a subject chose i over j and k over l. These covariances could be used in the off-diagonal elements of **V**.

Finally, we turn to *Maximum Likelihood* estimation of the unknowns. Here the goal is to pick the $\bar{s}_i$ and the σ_i^2 so as to maximize the likelihood of the sample. To begin, we define

$$f_{ij} = np_{ij}, \text{ that is, since}$$

$$p_{ij} = \frac{f_{ij}}{n},$$

i. e. the f_{ij} are the frequencies with which brand i is chosen over brand j. We also note that

$$p_{ji} = 1 - p_{ij} = \frac{n - f_{ij}}{n} \ .$$

We can now proceed to define the likelihood of the sample under the model as

$$\ell_0 = \prod_{i=1}^{t-1} \prod_{j=i+1}^{t} \hat{p}_{ij}^{f_{ij}} (1 - \hat{p}_{ij})^{n - f_{ij}} \ .$$

(12.22)

Note that with the two multiplication operators, the subscripts i and j run through each unique pair such that j > i. The log likelihood has its maximum at the same place as the likelihood. Taking logs on both sides leads to

$$\ln(\ell_0) = L_0 = \sum_{i=1}^{t-1} \sum_{j=i+1}^{t} f_{ij} \ln \hat{p}_{ij} + (n - f_{ij}) \ln(1 - \hat{p}_{ij})$$

(12.23)

which is much easier to deal with, being additive in form rather than multiplicative. Note here that we have used the rule of logarithms given in Equation (3.1), and also the rule from Equation (3.3).

The expression L_0 gives the log likelihood under the model, assuming that the model holds. The probability of the data under the general alternative that the pattern of frequencies is arbitrary is

$$\ell_A = \prod_{i=1}^{t-1} \prod_{j=i+1}^{t} p_{ij}^{f_{ij}} (1 - p_{ij})^{n-f_{ij}} . \tag{12.24}$$

Analogously to L_0, define L_A as $\ln(\ell_A)$. In that case

$$\hat{\chi}^2 = -2 \ln \frac{\ell_0}{\ell_A} = 2[L_A - L_0] . \tag{12.25}$$

Now we would need to figure out the derivatives of $\hat{\chi}^2$ with respect to each of the unknown parameters, the $\bar{s}_i$ and σ_i^2, and drive those derivatives to zero using nonlinear optimization as discussed in Section 3.9. When we reach that point we have our parameter estimates.
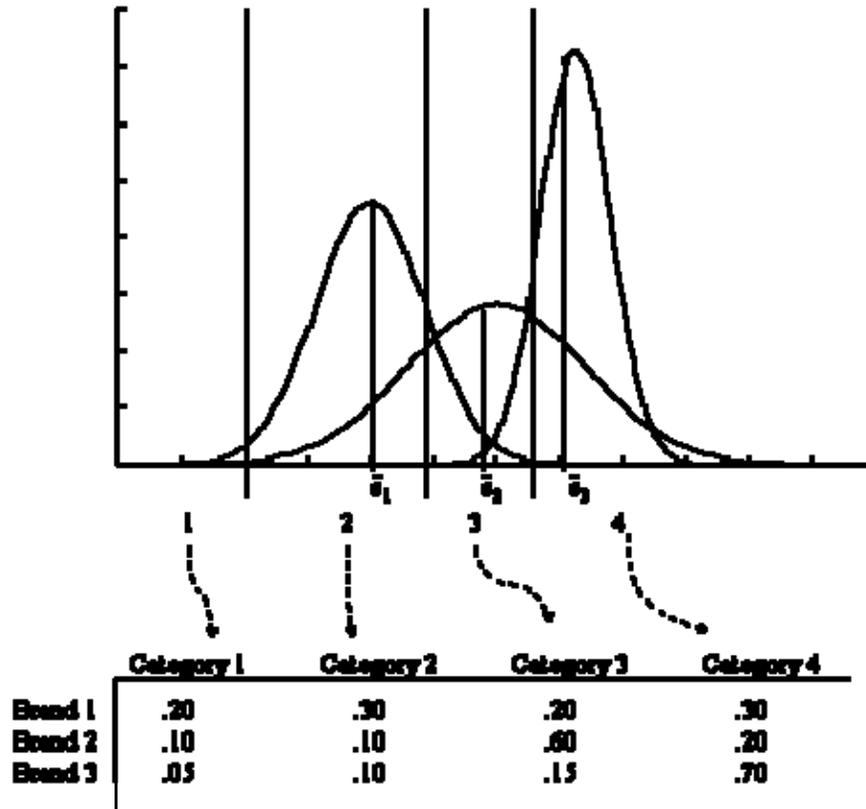
Note that for all of our estimation schemes; unweighted least squares, weighted least squares, modified minimum Chi Square, and Maximum Likelihood; we have q independent probabilities [t (t − 1) / 2] and 2 (t − 1) free parameters. The model therefore has q − 2 (t − 1) degrees of freedom.

12.5 *The Law of Categorical Judgment*

In addition to paired comparison data, Thurstone also contemplated absolute judgments, that is, when subjects assign ordered categories to objects without reference to other objects. For example, we might have a series of brands being rated on a scale as below,

<div align="center">

Like it a lot – Like it a little bit – Not crazy about it – Hate it
[ ]   [ ]   [ ]   [ ]

</div>

which is a simplified (and I hope marginally whimsical) version of the ubiquitous category rating scale used in thousands of marketing research projects a year. We assume that the psychological continuum is divided into four areas by three thresholds or cutoffs. In general, with a J point scale we would have J − 1 thresholds. We will begin with the probability that a subject uses category j for brand i. We can visualize our data as below:

	Category 1	Category 2	Category 3	Category 4
Brand 1	.20	.30	.20	.30
Brand 2	.10	.10	.60	.20
Brand 3	.05	.10	.15	.70

The probabilities shown above represent the probability that a particular brand is rated with a particular category. However, we need to cumulate those probabilities from left to right in order to have data for our model. The *cumulated probabilities* would look like the table below.

Brand 1	.20	.50	.70	1.00
Brand 2	.10	.20	.80	1.00
Brand 3	.05	.15	.30	1.00

Define the jth cutoff as c_j. We set $c_0 = -\infty$ and $c_J = +\infty$. We can then estimate values for c_1, c_2, $\cdots$, c_{J-1}. These cumulated probabilities are worthy to be called the p_{ij} and they represent the probability that brand i is judged in category j or less, which is to say, to the left of cutoff j. Our model is that each brand has a perceptual impact on the subject given by

$$s_i = \bar{s}_i + e_i \text{ with}$$

$$e_i \sim N(0, \sigma^2) .$$

In that case

$$\hat{p}_{ij} = \Pr[s_i < c_j] = \Pr[c_j - s_i > 0] . \tag{12.26}$$

But we have already seen a number of equations that look just like this! The probability that a normal random variable is greater than zero, Equation (12.8), has previously been used in the Law of Comparative Judgment. That probability, in this case of absolute judgment, is given by

$$\hat{p}_{ij} = \Phi\left[c_j - \bar{s}_i / \sigma_i\right]. \tag{12.27}$$

The importance of how to model categorical questionnaire items should be emphasized here. Such items are often used in factor analysis and structural equation models (Chapters 9, 10 and 11) under the assumption that the observed categorical ratings are normal. On the face of it, that would seem highly unlikely given that one of the assumptions of the normal distribution is that the variable is continuous and runs from $-\infty$ to $+\infty$! In the Law of Categorical Judgment, however, the variables s_i behave exactly that way. What's more, one can actually calculate the correlation between two Thurstone variables using what is known as a *polychoric correlation* and model those rather than Pearson type correlations.

12.6 *The Theory of Signal Detectability*

The final model to be covered in this chapter is another Thurstone-like model, but one invented long after Thurstone's 1927 paper. In World War II, Navy scientists began to study sonar signals, and more germane to marketing, they began to study the technician's response to sonar signals. Much later, models for human signal detection came to be applied to consumers trying to detect real ads that they had seen before, interspersed with distractor ads never shown to those consumers.

The theory of signal detectability (TSD) starts with the idea that a detection task has two distinct components. First, there is the actual sensory discrimination process, the resolving power if you will, of the human memory or the human senses being put to the test. This is related to our physiology, our sensitivity as receivers of the signal in question, and the signal-to-noise ratio. Second, there is a response decision involved. This is not so much a sensory issue as a cognitive one. It is related to bias, expectation, payoffs and losses, and motivation. For example, if you think you hear a submarine and it turns out you are wrong, the Captain may make you peel a crate of potatoes down in the mess hall. However, if you don't think that the sound you heard was a submarine and it turns out to have been one, you and the Captain will both find yourselves in Davy Jones' Locker, if you don't mind the nautical allusion. Given a particular ability to actually detect the sign of a sub, you might be biased towards making the first error and not making the second one. The TSD is designed to separate this response bias from your actual ability to detect subs.

Returning to our group of consumers being asked about ads they have seen, there are a number of ways to collect data. Assume that they have seen a set of ads. You are now showing them a series of ads which include the ads that they have seen along with some new ones that were never shown. Obviously, not including distractor ads is a little bit like giving a True/False test with no false items. You can ask them to say Yes or No; I saw that ad or I didn't. This is known as the *Yes/No Procedure*. You can also ask them on a ratings scale that might run from "Very Sure I Have Not Seen This Ad" on the left to "Very Sure I Have Seen This Ad" on the right. This is known as the *Ratings Procedure*. Finally, you can give them a sheet of paper with one previously exposed ad on it, and n - 1 other ads never before seen. Their task would be to pick the remembered ad from among the n alternatives, a procedure known as *n-alternative forced choice*, or n-afc for short. These procedures, and TSD, can be used for various sorts of judgments: Same/Different, Detect/No Detect, Old/New, and so forth. At this point, we will begin discussing the Yes/No task. The target ad that the consumer has seen will be called the signal, while the

distractor ads will be called the noise. We can summarize consumer response in the following table:

		Response	
		S	N
Reality	S	Hit	Miss
	N	False Alarm	Correct Rejection

Here the probability of a Hit plus the probability of a Miss sum to 1, as do the False Alarm and Correct Rejection rates. The consequence of a Yes/No trial is a value on the evidence continuum. The Subject must decide from which distribution it arose: the noise distribution or the distribution that includes the signal. We can picture the evidence distribution below.
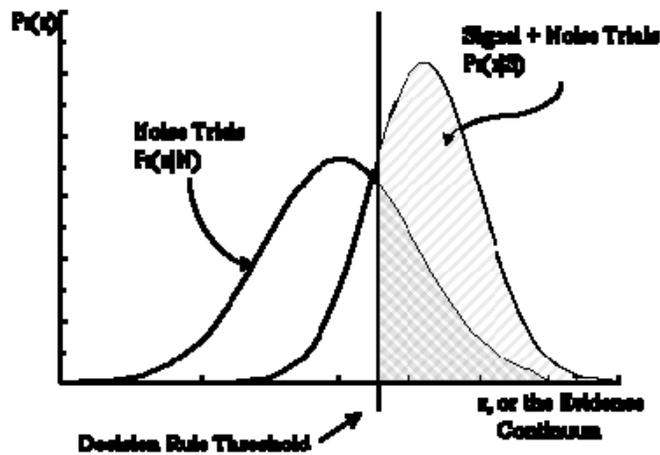
The x axis is the consumer's readout of the evidence to the consumer that the current trial contains an ad that they did indeed see. However, for whatever reason, due to the similarity between some target and some distractor ads, or other factors that could affect the consumer's memory, some of the distractor ads also invoke a relatively high degree of familiarity. The subject's task is difficult if the two distributions overlap, as they do in the figure. The difference in the means of the two distributions is called d′. The area to the right of the threshold for the Signal + Noise distribution, represented by lines angling from the lower left to the upper right, gives you the probability of a Hit, that is the Hite rate or HR. The area to the right of the Noise distribution gives you the False Alarm rate, or FAR. In the Figure, this is indicated by the double cross-hatched area. For noise trials we have

$$x_n = \overline{x} + e$$

and for signal + noise trials

$$x_s = \overline{x} + e + d'$$

where the parameter d′ represents the difference between the two distributions. We will assume that

$$e \sim N(0, \sigma^2)$$

and we fix $\bar{x} = 0$ and $\sigma^2 = 1$. Define the cutoff as c. Then

$$HR = Pr\,[Yes \mid Signal] = Pr(x_s > c) \tag{12.28}$$

$$= Pr\,(x_s - c > 0).$$

From Theorem (4.4) we can show that

$$E(x_s - c) = d' - c$$

and from Equation (4.8) that

$$V(x_s - c) = \sigma^2 = 1$$

so that

$$HR = \Phi(d' - c) \tag{12.29}$$

from Equation (12.8). As far as noise trials go,

$$FAR = Pr\,[Yes \mid Noise] = Pr(x_n > c) \tag{12.30}$$

$$= Pr\,[x_n - c > 0]\;.$$

Since

$$E(x_n - c) = -c$$

and

$$V(x_n - c) = \sigma^2 = 1,$$

we deduce that the

$$FAR = \Phi\,(-c)\;. \tag{12.31}$$

We can therefore transform our two independent data points, the HR and the FAR, into two TSD parameters, d′ and c. We can not test the model since we have as many parameters as independent data points. In order to improve upon this situation, we now turn to the Ratings procedure.

With ratings, we use confidence judgments to supplement the simple Yes/No decision of the consumer. The picture of what is going on under the ratings approach appears below;

	Very sure noise	noise	signal	Very sure signal
Signal	.30	.30	.20	.20
Noise	.20	.10	.40	.30
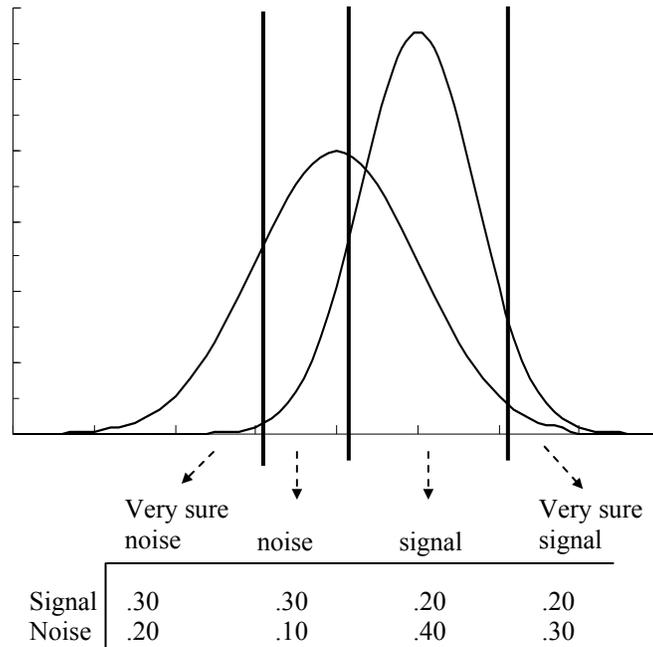
Just as we did in Section 12.5 with the Law of Categorical Judgment, we cumulate this table, which results in a stimulus-response table looking like

Signal	.30	.60	.80	1.00
Noise	.20	.30	.70	1.00

Each of the J-1 cutoffs, the c_j, defines a Hit Rate (HR_j) and a False Alarm Rate (FAR_j). Plotting them yields what is known as a Receiver Operating Characteristic, or ROC. Our pretend example is plotted below:
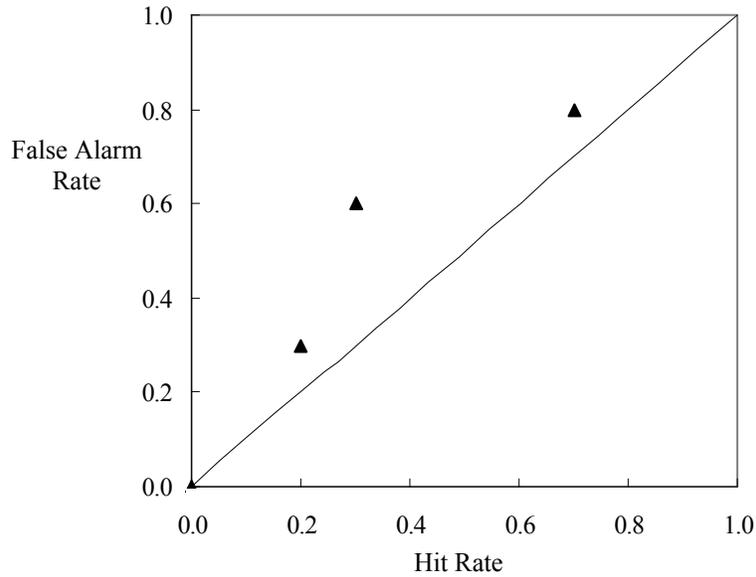
The shape of the ROC curve reveals the shape of the distributions of the signal and the noise. If we used z score coordinates instead of probabilities, the ROC should appear as a straight line. This suggests that we could fit the ROC using unweighted least squares. We will follow up on that idea shortly, but for now, let us review the model. For the Hit Rate for cutoff j we have

$$HR_j = Pr[x_s - c_j > 0]$$

$$= \Phi\left[(d' - c_j)/\sigma_s\right] \tag{12.32}$$

while for noise trials we have

$$FAR_j = Pr[x_n - c_j > 0]$$

$$= \Phi(-c_j) \tag{12.33}$$

Now we have $2 \cdot (J - 1)$ probabilities with only $J + 1$ parameters: $d', \sigma_s^2, c_1, c_2, \cdots, c_{J-1}$. Of course, we could use weighted least squares or maximum likelihood. Or we could plot the ROC using z scores and fit a line. In that case, the equation of the line would be
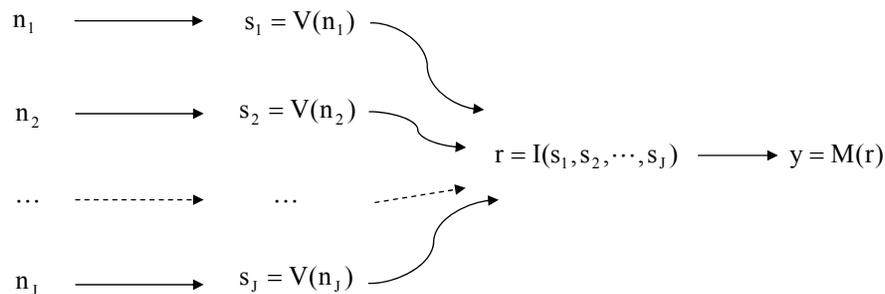
$$\hat{Z}_{HR} = d' + \frac{\sigma_n}{\sigma_s} Z_{FAR}$$

$$= d' + \frac{1}{\sigma_s} Z_{FAR} .$$

We close this chapter with just a word about the n-afc procedure. You can run this technique either sequentially or simultaneously. In either case, the consumer is instructed to pick exactly

one out of the n alternatives presented. There are no criteria or cutoffs in play in this procedure. According to the TSD, the percentage correct can be predicted from the area under the ROC curve.

12.7 *Functional Measurement*

We wrap up this chapter with a quick overview of what is known as functional measurement. We started the chapter talking about the relationship between the physical world and the mental impressions of that world. To round out the picture, after the sense impressions are transformed into internal stimuli, those stimuli may be combined, manipulated, evaluated, elaborated or integrated by the consumer into some sort of covert response. Then, this covert response is transformed into an observable behavior and voila, we have data to look at! A diagram will facilitate the explanation of the process:

$$n_1 \longrightarrow s_1 = V(n_1)$$
$$n_2 \longrightarrow s_2 = V(n_2)$$
$$\cdots \quad \cdots \qquad r = I(s_1, s_2, \cdots, s_J) \longrightarrow y = M(r)$$
$$n_J \longrightarrow s_J = V(n_J)$$

On the left, the inputs are transformed into mental events, we can call them discriminal values by the function $V(\cdot)$. In the case of physical input, $V(\cdot)$ is a psychophysical function. In the case of abstract input, we can think of $V(\cdot)$ as a valuation function. Then, the psychological or subjective values are integrated by some psychological process, call it $I(\cdot)$, to produce a psychological response. This might be a reaction to an expensive vacation package that goes to a desired location, or a sense of familiarity evoked by an ad. Finally, the psychomotor function $M(\cdot)$ transforms the mind's response into some overt act. This could be the action of putting an item in the shopping cart, or checking off a certain box of a certain questionnaire item. With the help of conjoint measurement, certain experimental outcomes allow all three functions to be ascertained.

References

Christoffersson, Anders (1975) Factor Analysis of Dichotomized Variables. *Psychometrika*. 40(1), 5-32.

Psychophysics

Batchelor, R.A. (1986) "The Psychophysics of Inflation," *Journal of Economic Psychology,* 7 (September), 269-90.

Monroe, Kent B. (1977) "Objective and Subjective Contextual Influences on Price Perception." In Arch G. Woodside, Jagdish Sheth, and Peter D. Bennet (Eds.) *Consumer and Industrial Buying Behavior.* NY: Elsevier-North Holland.

Sinn, Hans-Werner (1985) "Psychophysical Laws in Risk Theory," *Journal of Economic Psychology,* 6, 185-206.

Kamen, Joseph M. and Robert J. Toman (1970) "Psychophysics of Prices," *Journal of Marketing Research,* 7 (February), 27-35.

Gabor, Andre, Clive W. J. Granger and Anthony P. Sowter (1971) "Comments on 'Psychophysics of Prices',* Journal of Marketing Research,* 8 (May), 251-2.

Comparative Judgement

Thurstone, L.L. (1927) "A Law of Comparative Judgement," *Psychological Review*, 38, 368-89.

Freiden, Jon B. and Douglas S. Bible (1982) "The Home Purchase Process: Measurement of Evaluative Criteria through Pairwise Measures," *Journal of the Academy of Marketing Science,* 10 (Fall, 359-76

Signal Detection and Categorical Judgement

Cradit, J. Dennis., Armen Taschian and Charles Hofacker (1994) "Signal Detection Theory and Single Observation Designs: Methods and Indices for Advertising Recognition Testing," *Journal of Marketing Research,* (February), 117-27.

Parducci, Allen (1963) "Category Judgement: A Range-Frequency Model," *Psychological Review*, 72 (6), 407-18.

Tashchian, Armen J., J. Dennis White, and Sukgoo Pak (1988) "Signal Detection Analysis and Advertising Recognition: An Introduction Measurement and Interpretation Issues," *Journal of Marketing Research* 25 (November), 397-404.

Srinivasan, V. and Amiya K. Basu (1989) "The Metric Quality of Ordered Categorical Data," M*arketing Science*, 8 (Summer), 205-30.

Functional Measurement

Anderson, Norman H. (1982) "Cognitive Algebra and Social Psychophysics" in Bernd Wegener (Ed.) *Social Attitudes and Psychophysical Measurement,* Hillsdale, NJ: Lawrence Erlbaum.

Levin, Irwin P., Richard D. Johnson and Patricia J.Deldin (1985) "Framing Effects in Judgement Tasks with Varying Amounts of Information," *Organization Behavior and Human Decision Processes,* 36 (December), 362-77.

Levin, Irwin P. (1985) "How Changes in Price and Salary Affect Economic Satisfaction: Information Integration Models and Inference Processes," *Journal of Economic Psychology*, 6 (June), 143-55.

Johnson, Richard D. and Irwin P. Levin (1985) "More than Meets the Eye: The Effect of Missing Information of Purchase Evaluations," *Journal of Consumer Research,* 12 (September), 169-77.

White J. Dennis and Elise L. Truly (1989) "Price-Quality Integration and Warranty Evaluation: A Preliminary Test of Alternative Models f or Risk Assessment," *Journal of Business Research*, 19, 109-25.

Birnbaum, Michael H. (1982) "Controversies in Psychological Measurement," in Bernd Wegener (Ed.) *Social Attitudes and Psychophysical Measurement,* Hillsdale, NJ: Lawrence Erlbaum.

Lynch, John G., Jr. (1985) "Uniqueness Issues in the Decompositional Modeling of Multiattribute Overall Evaluations: An Information Integration Perspective," *Journal of Marketing Research*, 22 (February), 1-19