JOURNAL OF MATHEMATICAL PSYCHOLOGY: 5, 102–122 (1968)

Unidimensional Strength Theory and Component Analysis of Noise in Absolute and Comparative Judgments[1]

WAYNE A. WICKELGREN

Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

Using the decision rules and normal distribution assumptions of signal-detection theory as a base, a general strength theory of unidimensional absolute and comparative judgments is described in detail. The components of variance in both absolute and comparative judgments are considered, with particular emphasis on criterion variance in an absolute-judgment task and its relation to criterion variance in a comparative-judgment task. Some difficulties are noted in predicting comparative-judgment (forced-choice) probabilities from absolute-judgment ("yes–no") probabilities. The principal difficulties are concerned with the relative magnitudes of criterion variance in the two tasks, the correlation of distributions, and attention. The question of the equality of variances for different criteria (e.g., yes–no vs confidence criteria) is considered, and two methods are suggested for answering the question (one of which is a new type of operating characteristic). The notion of a random variable being a function of a real variable or being a function of another random variable is used to analyze the effects of noise in an independent variable on the distribution of a dependent random variable.

A strength theory is defined to be any theory in which the subject's choice of a response from a set of possible responses is determined with a probability of unity by the exact values of some number of random variables having continuous distribution functions on the real numbers. What makes a set of assumptions a strength theory is that the set of assumptions include a deterministic decision rule operating on one or more random variables. There need be no more to the theory than the decision rule with its specified input variables, which means that there need be very little theory relating the values of the psychological (intervening) variables to experimentally manipulable or measurable stimulus variables.

Thus, Thurstone's Law of Comparative Judgment (1927) is a strength theory with or without any laws relating the means and variances of the psychological dimensions involved in the judgment to the physical attributes of the stimuli. Hull's (1943) theories of animal learning and rote serial-list learning (Hull, Hovland, Ross,

[1] This work was supported by grant, MH 08890-03, from the National Institute of Mental Health, U.S. Public Health Service.

Hall, Perkins, and Smith, 1940) are also strength theories. Signal detection theory (Green, 1960; Green and Swets, 1966; Swets, Tanner, and Birdsall, 1961; Tanner and Swets, 1954) is a strength theory, as are the applications of this type of decision theory to other two-alternative sensory discrimination tasks (Creelman, 1962; Tanner, 1956; Weintraub and Hake, 1962), to perceptual recognition of 1 of $n > 2$ alternative stimuli (Green and Birdsall, 1958), to recognition memory for the amplitude or frequency of tones (Pollack, 1959; Wickelgren, 1966), and to verbal recognition memory (Egan, 1958; Norman and Wickelgren, 1965; Wickelgren and Norman, 1966).

Choice theory (Luce, 1959) is not a strength theory because a probabilistic decision rule is used to relate real-valued intervening variables to responses. A linear-operator stochastic learning model (Bush and Mosteller, 1955) is not a strength theory for many reasons, since changes in response probabilities are direct functions of independent variables without any intervening variables. Stimulus sampling theories (Atkinson and Estes, 1963) are not strength theories, though the reasons differ for component and pattern models. Component models (Estes, 1950) are not strength theories because they use a probabilistic decision rule to go from the proportion of conditioned stimulus elements sampled to the response. Pattern models (Estes, 1959), all finite-state learning models (Atkinson, Bower, and Crothers, 1965), buffer storage models (Atkinson and Shiffrin, 1965; Bower, 1964), two-state threshold models (Luce, 1963a), multi-state threshold models and neural quantum theory (Norman, 1964) are all not strength theories because the intervening variables (states) are discrete.

Although strength theories have been defined with sufficient generality to include theories postulating complex partitioning of multidimensional attribute spaces having any kind of multivariate probability density function defined over them, virtually all strength theories have been of two simple types. First, strength theories of "absolute judgments," such as single-interval signal detection and "yes-no" recognition memory, where a single intervening variable is assumed to be compared to a criterion (or cutoff) set somewhere on the scale of that intervening variable. This is easily extended to handle confidence judgments along with the "yes-no" judgment or to handle any set of responses that is ordered with respect to the values of the intervening variable. Second, theories of "comparative judgments," such as forced-choice tests of signal detection or recognition memory or a recall test of memory, where the decision rule is assumed to be to choose the response with the *greatest* value on the scale of a single intervening variable.

The present paper attempts to formulate a general version of strength theory applicable to both absolute and comparative judgments on a unidimensional scale. The primary emphasis of the paper is to analyze several theoretical components of the total noise in these two types of judgment tasks to determine how strength theories must handle noise from different sources both inside and outside the organism. Since the standard deviation of the noise in some condition provides the unit of measurement in a strength theory, this analysis is quite important in testing such a theory.

The present formulation of unidimensional strength theory borrows heavily from signal-detection theory for many of its basic concepts, and should be viewed essentially as a modification of the decision making aspects of signal-detection theory. However, much of signal-detection theory is missing, some of the basic concepts have been changed very slightly, and some theoretical problems that arise from these changes or that were implicit in previous formulations are explicitly considered.

What is missing is: (a) the interpretation of the dimension being judged as a likelihood ratio, (b) the theory of ideal observers, and (c) any substantive theory of sensory or memory systems that has been constructed on the decision making base of signal-detection theory.

The changes are as follows: (a) Criteria are assumed to have variances which are not insignificant in relation to the variances in the psychological representations of physical variables. (b) Variance in the physical variable itself is explicitly distinguished from variance introduced in the mapping of the physical variable into a psychological representation of it. (c) Attention and memory are assumed to affect both the psychological representation of a physical variable and the criterion, and these effects are assumed to be a function of the nature of the task. For example, the effects may be different for "yes-no" detection vs two-interval forced-choice. (d) Response bias terms for n-alternative multiple-choice tasks are handled in a manner completely analogous to the criterion term in a "yes–no" task, and this allows use of the simple maximum decision rule ("choose the alternative with the greatest value on the psychological scale"), instead of requiring a complex partitioning of an n-dimensional space (Luce, 1963b); Swets and Birdsall, 1956).

The main problem considered in this paper is how the assumption of substantial criterion variance affects the predictions of strength theory. The two specific predictions considered are the relationship between "yes-no" and rating operating characteristics and the predictions of comparative judgments (forced-choice tasks) from absolute judgments ("yes–no" or rating tasks).

UNIDIMENSIONAL STRENGTH THEORY

Let us consider a completely general strength theory of a subject making judgments of ordinal position on some physical dimension s', which we assume is represented by a single (real-valued) psychological dimension s. However, we assume that the mapping from s' to s involves noise, so that s' maps into a distribution of s values. Later on, we shall consider the effects of uncontrolled noise in the physical variable s', but for now we ignore it.

A random variable s, being a function of a real variable s', means that the parameters of the distribution function of the random variable are real-valued functions of a real variable. Also, in this paper, when two random variables are said to be equal, it will mean that they have the same distribution function.

To aid in understanding, it is probably helpful to keep one example in mind. Imagine that the subject is to detect a 1000 cps tone in noise, s' is the intensity of the 1000 cps tone, (s' may have only two values: $s' = s'_0 = 0$ and $s' = s'_1 > 0$), and s is the degree of activation of the internal representative of a 1000 cps tone. Notice, however, that the theory can be applied to a wide variety of different tasks. For example, s' could be any intensity dimension, and the two or more values of s' could be well above threshold with the subject asked to set a criterion for an absolute "loud–soft" judgment. Also, s' could be a qualitative dimension like frequency with subjects being asked to make "high-low" judgments. The theory can be applied to memory situations where s stands for strength in memory and s' takes on two values (presented or not presented) or many values (degree of recency or similarity to previously presented stimuli).

We may find that other independent variables, besides the one (s') the subject is asked to judge, affect the psychological dimension (s) on which the subject bases his judgment. For example, pitch judgments are affected by duration as well as frequency. Strength in short-term memory is affected by duration of original presentation, primacy, and similarity of test item to presented items, in addition to recency. This is easily represented by $s = (s', v'_1, ..., v'_m)$ and requires no significant extension of the theory as long as only one psychological dimension is used to make the judgment.

We shall consider 4 different types of judgment tasks, all involving judgments of s-ness: 2-alternative absolute judgment, n-alternative absolute judgment ($n > 2$), 2-alternative comparative judgment, and n-alternative comparative judgment ($n > 2$).

2—Alternative Absolute Judgment

In 2-alternative absolute judgments, subjects are assumed to set a single criterion c on the s dimension and make one response if $s - c \geqslant 0$ and the other response if $s - c < 0$. In our detection example, the responses might be R_1 = "present" and R_2 = "absent." Thus, the criterion decision rule for 2-alternative absolute judgment on a single dimension s is:

$$\text{Respond with } R_1 \text{, iff } t = s - c \geqslant 0 \text{,}$$
$$\text{Respond with } R_2 \text{, iff } t = s - c < 0 \text{,} \tag{1}$$

In signal-detection theory s is assumed to be a random variable and c a real variable, but it is most general to consider both s and c to be random variables. Certainly we know that criteria can vary considerably. At the same time one feels that, if the payoffs and subjective event probabilities remain constant, if subjects are instructed to maintain constant criteria, and if they are given practice, then criterion variance will be much reduced. However, the minimum criterion variance could still be large in relation to the variance in s (strength variance), especially if there is little variance in s'. The point is that we cannot answer this question, nor can we know if it makes any difference how we partition the total variance between s and c, unless we explicitly consider the possibility that there is variance in c, as well as s, and investigate the consequences of that assumption.

Strength theories almost always assume that random variables are normally distributed. This is particularly convenient when random variables are going to be added or subtracted because the resulting random variables are also normally distributed. Hence, we assume that $s \sim N(\mu_s, \sigma_s)$, $c \sim N(\mu_c, \sigma_c)$, from which it follows that $t = (s - c) \sim N(\mu_s - \mu_c, [\sigma_s^2 + \sigma_c^2]^{1/2})$.

Since we assumed that s was a function of the independent variable s', this means that $\mu_s = \mu_s(s')$ and possibly $\sigma_s = \sigma_s(s')$. However, we assume that c is not a function of s'.

Thus

$$p(R_1 \mid s') = \int_0^\infty N[\mu_s(s') - \mu_c, (\sigma_s^2(s') + \sigma_c^2)^{1/2}]. \tag{2}$$

Although $p(R_1 \mid s')$ is the directly measured dependent variable, it simplifies computation with this theory to transform probabilities into normal deviates such that the normal deviate corresponding to a probability is that which would be required to yield a probability of $p(R_1)$ in the right-hand tail of the normal distribution. Thus, we define the "tails normal deviate", TND(p), as

$$\mathrm{TND}(p) = \mathrm{TND} \int_a^\infty N(b, \sigma) = (b - a)/\sigma.$$

Applying the TND transform to Eq. 2 for two values of s', namely s_0' and s_1', we obtain:

$$\mathrm{TND}[p(R_1 \mid s_0')] = [\mu_s(s_0') - \mu_c]/[\sigma_s^2(s_0') + \sigma_c^2]^{1/2}, \tag{3}$$

$$\mathrm{TND}[p(R_1 \mid s_1')] = [\mu_s(s_1') - \mu_c]/[\sigma_s^2(s_1') + \sigma_c^2]^{1/2}. \tag{4}$$

The operating characteristic is obtained by solving the preceding two equations for μ_c and equating namely,

$$\mathrm{TND}[p(R_1 \mid s_1')] = \{[\sigma_s^2(s_0') + \sigma_c^2]^{1/2}/[\sigma_s^2(s_1') + \sigma_c^2]^{1/2}\}\, \mathrm{TND}[p(R_1 \mid s_0')] + [\mu_s(s_1') - \mu_s(s_0')]/[\sigma_s^2(s_1') + \sigma_c^2]^{1/2}. \tag{5}$$

If σ_c is assumed constant for different values of μ_c, then such an operating characteristic will be linear with a slope of $[\sigma_s^2(s_0') + \sigma_c^2]^{1/2}/[\sigma_s^2(s_1') + \sigma_c^2]^{1/2}$ and a y-intercept of $[\mu_s(s_1') - \mu_s(s_0')]/[\sigma_s^2(s_1') + \sigma_c^2]^{1/2}$, when we plot the TND's of the two response probabilities against each other. If in different sessions we can induce the subject to adopt different mean criteria, μ_c, (by manipulating instructions, payoffs, or *a priori* presentation probabilities) with no change in the standard deviation of the criterion, σ_c, then we can obtain a number of distinct points of this operating characteristic and can test the normal distribution assumption by determining how well the points are fitted by a straight line. Assuming we are satisfied by their approximation to a straight line, we can obtain empirical estimates of $[\sigma_s^2(s_0') + \sigma_c^2]^{1/2}/[\sigma_s^2(s_1') + \sigma_c^2]^{1/2}$ and $[\mu_s(s_1') - \mu_s(s_0')]/[\sigma_s^2(s_1') + \sigma_c^2]^{1/2}$ by the slope and y-intercept of the best-fitting straight line through these points. Actually, the d' of signal-detection theory is the absolute value of the x-intercept, namely,

$$\begin{aligned} d' &= |\mu_s(s_0') - \mu_s(s_1')|/[\sigma_s^2(s_0') + \sigma_c^2]^{1/2} \\ &= [\mu_s(s_1') - \mu_s(s_0')]/[\sigma_s^2(s_0') + \sigma_c^2]^{1/2}. \end{aligned} \tag{6}$$

Often a less variable estimate of the distance between $\mu_s(s_1')$ and $\mu_s(s_0')$ is given by twice the value of the y-coordinate of the point of intersection of the operating characteristic with the negative diagonal, namely,

$$d_s = 2\,\text{TND}[p(R_1 \mid s_1')] = \frac{[\mu_s(s_1') - \mu_s(s_0')]}{1/2\{[\sigma_s^2(s_0') + \sigma_c^2]^{1/2} + [\sigma_s^2(s_1') + \sigma_c^2]^{1/2}\}}. \tag{7}$$

The proof of the above follows by simple algebra from Eq. 5 by noting that the x and y coordinates of the intersection with the negative diagonal are $\text{TND}[1 - p(R_1 \mid s_1')]$ and $\text{TND}[p(R_1 \mid s_1')]$, respectively, and that $\text{TND}[1 - p(R_1 \mid s_1')] = -\text{TND}[p(R_1 \mid s_1')]$.

When the slope of the operating characteristic is unity, $d_s = d'$. Since the estimated slopes of operating characteristics often demonstrate considerable *random* variation, it is often preferable to use d_s, rather than d'. However, d_s is measured in units of the average of two standard deviations. Thus, if slope varies *systematically* with d', d' is to be preferred to d_s. The reason is that the unit of measurement for d' is constant for all operating characteristics that are obtained by plotting TND's of the response probabilities for many different conditions on the y-axis against the TND of the response probability for a single condition on the x-axis. On the other hand, the unit of measurement for d_s would be different for every such operating characteristic.

Notice that when criterion variance is considered, it is clear that the slope of the operating characteristic does *not* provide a measure of the ratio of the standard deviations of the two s-distributions, as has previously been assumed. One obtains a measure of $[\sigma_s^2(s_0') + \sigma_c^2]^{1/2}/[\sigma_s^2(s_1') + \sigma_c^2]^{1/2}$, which requires one to estimate the relative size of the variance of some strength distribution in relation to criterion variance before one can determine how fast the strength variance is increasing with the mean. The conclusion of Nachmias and Steinman (1963) and Swets *et al.* (1961) that the standard deviation increases at a rate of .25 times the increase in the mean of the strength distributions in visual signal detection, is valid only if criterion variance is negligible in relation to strength variance. Failure to consider the effects of criterion variance on the slope parameter provides one explanation for why Markowitz and

Swets (1967) found a higher "mean to sigma" ratio for more intense auditory signals than for less intense signals. In fact, one can assume that the mean to sigma ratio for strength distributions should be a constant and use this to estimate the relation between criterion variance and strength variance and at the same time estimate the true mean to sigma ratio.

N—Alternative Absolute Judgment

Subjects can be asked to provide more information about their s value in response to a particular s' value by making any one of n responses, each of which is assumed to correspond to a region on the s dimension between two criteria. If there are n or more value of s', the interpretation of an n-alternative absolute judgment task is quite obvious. The subject is being asked to use his s dimension to identify more or less exactly the value of the stimulus on the s' dimension. If there are only two values on the s' dimension (present, absent), the subject can still be asked to give a reasonable n-alternative analysis of the s dimension by eliciting confidence (rating) judgments from the subject in addition to his "yes–no" response: "yes" responses with high confidence being assumed to correspond to very high values of s, "yes" responses with low confidence to somewhat lower values of s, and "no" responses with high confidence to the lowest values of s. Whichever interpretation is given to the subject for the rating responses, it is reasonable to assume that the responses are ordered with respect to the s dimension, provided that the psychological dimension s exists and is the basis for the judgment.

To aid in visualizing strength theory for n-alternative absolute judgments, assume that there is no variation in the criteria. In that case, Fig. 1 provides an accurate picture of what is assumed by strength theory for 6-alternative judgments. If there is normal variation in the criteria, then one can visualize the criteria as whirring back and forth around their means.

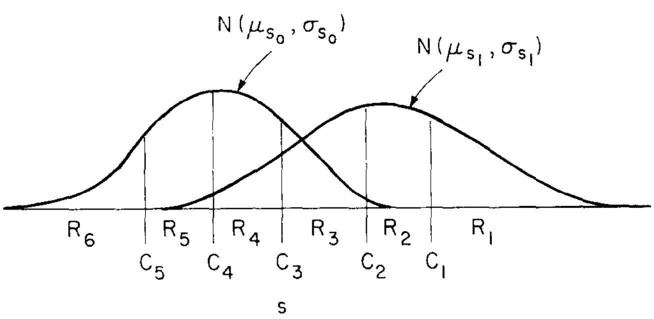

FIG. 1. Strength theory for n-alternative absolute judgments with 2 values of s', s'_0, and s'_1, and no variation in the criteria, c_i.

The algebraic development of the theory for n-alternative absolute judgments is completely identical to that for 2-alternative judgments, except that one has as many $s - c_i$ terms as there are criteria, c_i, and for each criterion c_i one lumps all responses to the right of $c_i(R_1, ..., R_i)$ considering them as one response $(R_{x<i})$ and all the responses to the left of c_i as the other response. Having done this, all of the theory developed for 2-alternative judgments applies to n-alternative judgments. Furthermore, this method of n-alternative judgments is another way of generating several distinct points on an operating characteristic, as each distinct criterion c_i yields another distinct point. Applications of this rating method of generating operating characteristics for signal detection are to be found in Egan, Schulman, and Greenberg (1959) and Swets *et al.* (1955, 1961) and for recognition memory in Egan (1958).

Validity of Strength Theory for Absolute Judgments

It should be noted that, for both binary decision and rating tasks, the prediction of a linear operating characteristic on normal–normal coordinates and the preceding interpretations of its slope and intercept parameters depend on a number of assumptions, besides the criterion decision rule. First, the variance of the criteria must be constant, independent of mean criterion location. Second, the experimental conditions (payoffs, *a priori* probabilities, or instructions) that alter mean criterion location must have no effect on the strength distributions. Third, the strength and criterion distributions must be normal. If the first two assumptions are valid, then the normal distribution assumption is directly testable by the shape of the operating characteristic on normal–normal coordinates, namely, it should be a straight line.

The second assumption seems automatically valid for a rating task, since the subject is setting all the different criteria simultaneously and, thus, the strength distributions cannot be different for the different mean criterion placements. The subject's attentional set and all other potentially variable aspects of the sensory or memory system under study must be in the same state for each criterion on each trial of a rating task. The distribution of these properties over many trials must, therefore, also be the same for each criterion in a rating task.

This is not necessarily so for the various binary decision methods. In fact, one interpretation of the data of Markowitz and Swets (1967) and Schulman and Greenberg (1960) is that varying *a priori* probabilities can alter strength distributions. Furthermore, the discrepancy between the slopes for operating characteristics obtained by binary decisions and by ratings in Swets *et al.* (1961) may indicate that varying payoffs can also affect strength distributions. Alternatively, the data of these three studies may result from changes in criterion variance as a function of *a priori* probabilities or payoffs. Whichever interpretation is correct, these studies indicate that manipulating *a priori* probabilities or payoffs is not the ideal method for generating operating characteristics.

On the other hand, Egan *et al.* (1959) found good agreement between operating characteristics obtained by binary decisions and by ratings, when the criterion in the binary decision task was manipulated by instructions to maintain a "strict," "medium," or "lax" criterion (as defined by the desired frequency of "yes" responses in different conditions). Thus, the existing data suggest that instructional manipulation of the desired frequency of "yes" responses is as effective as the rating method in avoiding any differential effects of different mean criterion placements on the strength distributions.

The assumption of equal criterion variance for all mean criterion placements poses an interesting theoretical problem, to which one can take several different approaches. In the first place, it should be understood that, at present, there is no obviously best unit of measurement for any ordered psychological *s*-dimension. We have a choice, though some choices may be better than others in the sense of allowing simpler theories to fit the data. In many ways the most natural choice of a unit is the standard deviation of the criterion variability *at each point along the s-dimension*, where the different mean criterion placements are obtained by a single experimental method. Notice that, from this point of view, it is meaningless to ask if the variances of the different mean criterion placements are equal. They are equal by definition. However, if we have some other method of varying mean criterion location, we can determine whether the two methods yield the same criterion variabilities at each point by determining whether the operating characteristics they produce are identical. If they are not identical, and if we feel sure that the different methods do not differentially affect the strength distributions, then a different d' value tells us that the lower operating characteristic has a greater criterion variance (at least in the vicinity of the mean of the noise distribution). A different slope value tells us that the ratio of the criterion variances at different points is different.

Of course, there is no guarantee that with any method of producing different mean criterion locations, the spacing of values on any given *s*-dimension will be such as to yield strength distributions of the same simple (e.g., normal) form with means and variances that vary in a simple way with the values of certain physical variables. To work on a problem, one assumes that there is some way of spacing the values on a particular *s*-dimension such that the strength distributions are of the same simple (normal) form. Given that assumption, one rejects any method that produces "ugly" results on the grounds that the different criteria have different variances (measured on this unknown "ideal" scale) or on the grounds that the mean criterion placement affects the strength distributions, whichever seems more plausible. One hopes that one or more methods of varying criteria will be found which yield "pretty" results.

On the basis of present evidence, it would seem that varying *a priori* probabilities or payoffs is not ideal, because these methods may violate the assumption that the strength distributions are independent of the mean criterion placement or the assumption that criterion variance is independent of mean criterion location. However, this conclusion is far from being definitely established. With the proper feedback, with the proper instructions, or with the frequent interpolation of trials where the subject knows in advance what is being presented, it might well be that the payoff or *a priori* probability method would not violate the assumptions that the strength distributions are independent of the mean criterion placements or that criterion variance is independent of mean criterion location.

Although the rating method has the advantage of insuring that the different criteria are "looking" at the same strength distributions, it has the possible disadvantage that there may be increased criterion variability due to the larger number of criteria that must be maintained simultaneously. However, the mere fact that the 9 criteria for a 10-category rating are more variable than the 1 criterion for a binary response is no argument against use of the 10-category rating scale. Inches are not a poorer unit of length measurement than "finger widths" because the number of inches in a given distance will (generally) be less than the number of "finger widths." The key issue is whether each of the 9 different criteria has the same variance on the unknown "ideal" scale.

Whenever the rating categories are given the confidence judgment interpretation, there is almost always an explicit separation of the response categories into two classes corresponding to the two categories of the analogous binary response method. Thus, there is the possibility that the "yes–no" criterion in the confidence judgment method may have a lower variance than the other "confidence" criteria. Whenever this is the case, the operating characteristic obtained by the rating method will have a "peak" at the "yes–no" point. A statistical test of the criterion variance for any given point compared to the average criterion variance of the other points of an operating characteristic can be obtained by drawing a best-fitting straight line through the other points and observing whether the point one is testing lies above or below the line. This must be done on a reasonable number of different operating characteristics to determine if there is any systematic tendency for the point corresponding to a given criterion (such as the "yes–no" criterion) to lie above or below the best-fitting line through the other points.

A quick check of a few studies that have employed the confidence judgment method to obtain operating characteristics reveals some that show no significant tendency for the "yes–no" point to lie above the operating characteristic drawn through the other points (Egan *et al.*, 1959; Swets *et al.*, 1961), some that show very slight, but statistically significant, tendencies for the "yes–no" point to lie above the line through the other points (Egan, Greenberg, and Schulman, 1961; Markowitz and Swets, 1967; Schulman and Mitchell, 1966; Wickelgren and Norman, 1966), and some that show slightly larger deviations (Watson, Rilling, and Bourbon, 1964). There is some tendency for the larger deviations to occur in studies with a larger number of response categories. For example, the Watson *et al.* (1964) study used a mechanical sliding scale, which was divided into 36 categories by the experimenter, but which was a continuous scale to the subject (except that the center was clearly marked).

Another example of somewhat larger discrepancies between the variance of the "yes–no" criterion and the variance of "confidence" criteria occurred in an unpublished experiment Don Norman and I did, using latency in a recognition memory task as a measure of lack of confidence. Interpreting latency as a type of confidence measure means assuming that the responses (decision–latency pairs) in the experiment are ordered along the s-dimension from high to low s values as follows: short-latency "yes" responses, long-latency "yes" responses (lumping together all latencies longer than some value), long-latency "no" responses, and finally short-latency "no" responses. The criterion separating the longest latency "yes" response from the longest latency "no" response is simply the "yes–no" criterion with a variance independent of the variance in the relationship between latency and $s - c$ value. However, the other criteria include various degrees of variation due to the apparently variable relationship between latency and $s - c$ value, and this variance combines with the "yes–no" criterion variance to yield a larger variance for the other criteria than for the "yes–no" criterion. This lesser variance of the "yes–no" criterion produced a peak in the operating characteristic at that point.

What conclusions can be drawn about the variability of different criteria using the rating method? Apparently the "yes–no" criterion may have slightly lower variance than the confidence criteria under some circumstances. Non-verbal continuous scales, such as the mechanical sliding scale and latency, probably give the greatest discrepancy. Verbal rating scales with 10 or less categories appear to yield very small discrepancies, which can probably be ignored. However, if the "yes–no" point is consistently elevated in an experiment, it might be wise to delete it completely and use only the confidence points to determine the operating characteristic.

Many other questions about rating scales remain to be answered. Ignoring any peak at the "yes–no" point, how does criterion variability increase with an increasing number of rating categories? Can the variability of different criteria be more exactly equated by instructions to use categories equally often and by practice accompanied by feedback as to the frequency with which categories are being used? How do operating characteristics obtained with unequal numbers of categories on each side of the "yes–no" criterion compare with operating characteristics obtained from the usual symmetric rating scales? Do rating scales with a central "don't know" category yield different operating characteristics from rating scales which force a "yes–no" decision? What happens when instructional manipulation of the frequency of "yes–no" responses is combined with the rating method to yield one binary-response operating characteristic and rating operating characteristics around each point of the binary operating characteristic? If subjects make an n-category choice and then are sometimes (and sometimes not) asked to locate their response more precisely within the chosen category, is any information transmitted by these secondary ratings and where do the points fall in relation to the operating characteristic based on the primary ratings?

Criterion Operating Characteristic

If and only if one has strong reasons to believe that the variance of the underlying strength distributions is constant, independent of the mean (measured on the unknown "ideal" scale), then there is an elegant way to compare systematically the variability of two different criteria by examining the slope of a new type of operating characteristic. This operating characteristic is obtained from analogues to Eqs. 3 and 4 which assume the same value of s' and two criteria, c_i and c_j, namely:

$$\text{TND}[p(R_1 \mid c_i)] = [\mu_s - \mu_{c_i}]/[\sigma_s^2 + \sigma_{c_i}^2]^{1/2},$$
$$\text{TND}[p(R_1 \mid c_j)] = [\mu_s - \mu_{c_j}]/[\sigma_s^2 + \sigma_{c_j}^2]^{1/2}.$$

Assuming that σ_s is constant for all μ_s, we can solve these equations for μ_s and equate, obtaining a "criterion operating characteristic" which is linear on normal–normal probability coordinates, with a slope of $[\sigma_s^2 + \sigma_{c_i}^2]^{1/2}/[\sigma_s^2 + \sigma_{c_j}^2]^{1/2}$.

$$\text{TND}[p(R_1 \mid c_j)] = [\sigma_s^2 + \sigma_{c_i}^2]^{1/2}/[\sigma_s^2 + \sigma_{c_j}^2]^{1/2}\, \text{TND}[p(R_1 \mid c_i)] + (\mu_{c_i} - \mu_{c_j})/(\sigma_s^2 + \sigma_{c_j}^2)^{1/2}. \qquad (8)$$

If c_i and c_j have equal standard deviations, $\sigma_{c_i} = \sigma_{c_j}$, then the operating characteristic represented by Eq. 8 will have a slope of unity. The d' value for such an operating characteristic represents the distance between the two criteria in standard deviation units.

This criterion operating characteristic can be applied equally well to the comparison of any two criteria obtained by the rating method or by some binary response method. When the rating method is used, $R_{x<i}$ can be substituted for $(R_1 \mid c_i)$ in Eq. 8. However, to obtain the criterion operating characteristic it is necessary to have the *same* pair of criteria being applied to a wide variety of strength distributions. Thus, in a signal detection task, it would be necessary to present all the different levels of signal intensity intermixed in each session of the experiment in unknown, random order. It is not appropriate for the purpose of this analysis to plot on the same criterion operating characteristic, points obtained from sessions with different sets of possible signal intensities, because a subject might well alter the distance between two criteria when the stimulus ensemble is altered.

2 — Alternative Comparative Judgment

The general strength theory for this type of task derives from Thurstone's (1927) Law of Comparative Judgment and the decision aspects of signal detection theory's analysis of 2-alternative forced-choice (Tanner and Swets, 1954), as extended by Luce (1963b) to include response biases. The subject is presented with two values of s', namely s'_0 and s'_1 and asked to choose the greater. Were there no possibility of response bias, we would assume he performs this task by selecting the maximum of $s_0 = s(s'_0)$ and $s_1 = s(s'_1)$ (maximum rule). For 2-alternative comparative judgments, it is mathematically equivalent to assume the subject bases his decision on the difference between s_0 and s_1 (difference rule).

However, it is generally necessary to distinguish between s'_0 and s'_1, in temporal or spatial position, in addition to s' value, yielding four different stimuli, s'_{01}, s'_{02}, s'_{11}, s'_{12}. It is possible that this distinction could affect the judgment. Some of these effects can be called response biases, since the perceptible spatial or temporal distinction between the two stimuli is the basis of the verbal response allowing the subject to communicate to the experimenter which stimulus he thinks has the higher s' value. There are at least two different ways to handle these response biases, only one of which generalizes to more than two alternatives.

The one which does not generalize is to use the difference rule modified so that when the subject subtracts the second stimulus' s value from the first stimulus' s value, the difference must be greater than b, which is not necessarily equal to zero.

The way which generalizes to more than two alternatives is to use the maximum rule and incorporate the bias variables into the functions $t_{0k} = t(s'_0, k)$, $t_{1k} = t(s'_1, k)$, where $k = 1$ or 2 depending on whether the stimulus was presented first or second. If $t_{ik} = t(s'_i, k) = s(s'_i) - c_k = s_i - c_k$, then the maximum rule with bias is equivalent to the difference rule with bias.

Note, however, that the interpretation of the stage at which the bias term enters into the process is different for the maximum rule and the difference rule. For the maximum rule, the bias terms are essentially the same as criteria in the absolute judgment task, and the criterion is subtracted from each s term before the two $s - c$ terms are compared. This means that there are really two criteria involved in the 2-alternative comparative judgment task, whereas the more familiar interpretation has been that of a single criterion on the difference between the two s terms. For two alternatives, the probability that $s_i - c_1 > s_j - c_2$ is identical to the probability that $s_i - s_j - b > 0$, where $b = c_1 - c_2$. Thus, both the difference rule and the maximum rule are described by Eq. 9.

$$\begin{aligned} &\text{Respond } R_1(\text{“first } s' \text{ greater”}) \text{ iff } s_i - s_j - b > 0, \\ &\text{Respond } R_2 (\text{“second } s' \text{ greater”}) \text{ otherwise.} \end{aligned} \tag{9}$$

So, in one sense, it makes no difference whether one thinks of there being two criteria, one applied to each s term, or one criterion applied to the difference between the s terms. But, in another sense, it does matter, because it changes the most natural assumption concerning the total criterion variance in the comparative-judgment task in relation to the criterion variance in the absolute-judgment task. If there are really two criteria and $b = c_1 - c_2$, then the most natural assumption is that

$\sigma_{c_1}^2 = \sigma_{c_2}^2 = \frac{1}{2}\sigma_b^2 = \sigma_c^2$. On the other hand, if there is only a single criterion, then the most natural assumption is that $\sigma_b^2 = \sigma_c^2$.

Usually we want to assume that all the variables in Eq. 9 are normally distributed. Also, we again must assume that σ_b is constant with changing criterion mean. In this case, operating characteristics are easily derived by plotting TND$[p(R_1 \mid s_{11}', s_{02}')]$ against TND$[p(R_1 \mid s_{01}', s_{12}')]$, yielding:

$$d_2' = 2[\mu_s(s_1') - \mu_s(s_0')]/[\sigma_s^2(s_1') + \sigma_s^2(s_0') - 2r\sigma_s(s_1')\,\sigma_s(s_0') + \sigma_b^2]^{1/2},$$

where r is the correlation between the two distributions.

If $r = 0$, $\frac{1}{2}\sigma_b^2 = \sigma_c^2$ (the two-criteria interpretation), and the absolute-judgment operating characteristic has unit slope (i.e., $\sigma_s^2(s_1') = \sigma_s^2(s_0')$), then $d_2' = \sqrt{2}\,d'$, and one can predict the 2-alternative comparative-judgment d_2' from the absolute-judgment d'. On the other hand, if *any* of these conditions is grossly violated, it is not possible to predict d_2' from d' without estimating additional parameters. Note that the one-criterion interpretation for comparative judgment does not predict that $d_2' = \sqrt{2}\,d'$.

In view of all the conditions that must hold to predict the "forced-choice" d_2' strictly from the "yes–no" d' (without using the slope of the "yes–no" operating characteristic or estimating σ_b and r from the forced-choice data), it is quite amazing that the predicted d_2's have not differed systematically from obtained d_2's in several studies of signal detection (Tanner and Norman, 1954; Tanner and Swets, 1954; Shipley, 1965; Swets, 1959). The present analysis makes it abundantly clear that a strength theory may be correct without d_2' being equal to $\sqrt{2}\,d'$.

To see just how amazing it is that there has been no systematic deviation from the prediction that $d_2' = \sqrt{2}\,d'$, let us examine the effects of deviation from each of the three assumptions that are necessary for $d_2' = \sqrt{2}\,d'$ to hold.

An example of the effects of correlation between the two strength distributions is easily obtained for the case where $\sigma_s^2(s_0') = \sigma_s^2(s_1') = \sigma_c^2 = \frac{1}{2}\sigma_b^2$. In this case, $d_2' = \sqrt{2}\,d'[1/(1 - r/2)]^{1/2}$. The maximum value of d_2', obtained for $r = 1$, is 41% greater than $\sqrt{2}\,d'$, but a more reasonable correlation of .5 would produce a value of d_2' only 15% greater than $\sqrt{2}\,d'$. Still, such effects are not negligible, and the absence of systematic deviations from the $\sqrt{2}\,d'$ prediction suggests either that correlations are not too far from zero, or that the effects of nonzero correlation are being counteracted by other effects.

An example of the effects of deviation from the assumption that $\sigma_b^2 = 2\sigma_c^2$ is easily obtained for the case where $\sigma_s^2(s_0') = \sigma_s^2(s_1') = \sigma_c^2$ and $r = 0$. The general form of the relationship between σ_b and σ_c can be given by $\sigma_b^2 = 2w\sigma_c^2$, where $w = 1$ is the assumption required for the $\sqrt{2}\,d'$ prediction. For the case considered here, $d_2' = \sqrt{2}\,d'[2/(1 + w)]^{1/2}$. Since w can have any value greater than zero, the possible effects of deviation from the assumption that $w = 1$ are d_2' values anywhere from 100%

less to 41% greater than $\sqrt{2}\, d'$. The single criterion assumption for 2-alternative comparative judgment ($w = .5$) would yield a value of d_2' that was 33% greater than $\sqrt{2}\, d'$.

The effects of deviations from the assumption that the absolute judgment operating characteristic has unit slope are easily obtained for the case where $\sigma_c^2 = \sigma_b^2 = 0$ and $r = 0$. Let us represent the general case of nonunit slope by $\sigma_s^2(s_1) = m^2\sigma_s^2(s_0)$, where m is the reciprocal of the slope of the absolute judgment operating characteristic. For this case, $d_2' = \sqrt{2}\, d'[2/(1 + m^2)]^{1/2}$. Thus, a slope of .5 for the absolute judgment operating characteristic would produce a value of d_2' that was 37% less than the $\sqrt{2}\, d'$ prediction.

A recent study of the relationship between absolute and 2-alternative comparative judgment in auditory signal detection by Schulman and Mitchell (1966) eliminated any effects of nonunit slope in the absolute judgment operating characteristic by using a statistic of the operating characteristic in both "yes–no" and forced choice, D_{YN} and D_{FC}, for which the relationship $D_{FC} = \sqrt{2}\, D_{YN}$ holds, irrespective of the slope of the "yes–no" operating characteristic. That statistic is the perpendicular distance from the origin to the operating characteristic on a normal–normal plot. When the effects of slope were thus eliminated, Schulman and Mitchell's results still showed no systematic deviation from the $\sqrt{2}\, d'$ prediction, but there was enough unsystematic deviation to suggest the possibility of competing deviations, that sometimes produced overpredictions and sometimes produced underpredictions.

Assuming that one has taken the slope of the absolute-judgment operating characteristic into account, one still has to worry about the effects of nonzero correlation between the two strength distributions and the relationship between the criterion variances in the two tasks. However, there is another possible complication. The level of attention to each of two alternatives may not be as high as that for a single alternative. If attention affects the strength distributions, as it well may, then the difference between the means of the two strength distributions may be reduced for 2-alternative comparative judgment and the variances may also be affected.

One can imagine that the effects of attention might be present both in cases where the 2-alternatives are two successive intervals either of which may contain a signal or two alternative stimuli that could be presented in a single interval (e.g., either a light or a sound). In the successive case, the issue is whether attention can be sustained at the same high level over two successive intervals as over one interval. In the simultaneous case, the issue is whether one can attend to two stimulus channels as well as to a single channel.

Furthermore, not only may level of attentional set be lower for two alternatives than for one, but it may be greater for one of the two alternatives than for the other. In addition to this, in the successive alternative case, there could be sequential effects of the nature of the stimulus presented in the first interval on the level of attention or state of adaptation in the second interval.

All these attentional and adaptational effects can be studied by experiments in which both or neither of the "alternative" stimuli may be presented, as well as one or the other. In such an experiment the subject makes an absolute judgment about the presence or absence of each stimulus. Ordinary absolute and 2-alternative comparative judgments tasks should also be performed by the same subject with the same stimuli, and the pattern of the results for all conditions will show whether attention and adaptation are influencing the results in the ordinary prediction of comparative judgment from absolute judgment.

When one considers all the ways in which the $\sqrt{2}\, d'$ prediction might fail for reasons that have nothing to do with the essential validity of strength theory for both absolute and comparative judgments, it is truly amazing that it has not failed thus far. However, the present analysis makes it clear that, if the $\sqrt{2}\, d'$ prediction fails in some future application of strength theory, one cannot reject strength theory without a detailed study of the reasons for the failure.

N-Alternative Comparative Judgment

The biased maximum rule generalizes easily from 2-alternative comparative judgment to n-alternative comparative judgment to yield:

$$\text{Respond } R_k(k\text{th stimulus is maximum}), \text{ iff } t_{ik} \geqslant t_{jm} \text{ for all } m \neq k,$$

$$m = 1, \ldots, n, \text{ where } t_{ik} = s_i - c_k \text{ and } t_{jm} = s_j - c_m .$$

Assuming that $s_i \sim N[\mu_s(s_i'), \sigma_s(s_i')]$ and $c_k \sim N[\mu_{c_k}, \sigma_{c_k}]$, then the probability density function for t_{ik} is $f_{ik}(t) = N[\mu_s(s_i') - \mu_{c_k}, (\sigma_s^2(s_i') + \sigma_{c_k}^2)^{1/2}]$. Let the cumulative distribution function for t_{jm} be $F_{jm}(t) = \int_{-\infty}^{t} f_{jm}(x)\, dx$. Now if we assume that all the random variables are independent of each other, the probability of responding R_k is given by:

$$p(R_k) = \int_{-\infty}^{\infty} f_{ik}(t) \prod_{\substack{m=1 \\ m \neq k}}^{n} F_{jm}(t)\, dt.$$

If $\mu_{c_k} = \mu_{c_m}$ and $\sigma_{c_k} = \sigma_{c_m} = \sigma_c$ for all k and m, then the above equation can be solved numerically given the d' and slope parameters of enough "yes–no" operating characteristics to involve all the s_i' conditions at least once. Tables predicting $p(R_k)$ from a single "yes–no" d' value are given by Elliott (1964) for the case where $f_{1k}(t) = N[C_1 + d', C_2]$ and $f_{0m}(t) = f_{0v}(t) = N[C_1, C_2]$ for all $(m, v) = 1, \ldots, n$ and $(m, v) \neq k$.

To my knowledge, there is no satisfactory way to analyze n-alternative comparative judgments for cases where $c_k \neq c_m$ for some k and m or where the various strength distributions for the n-alternatives are correlated. Cases where the strength distributions are independent, normal, and $c_k = c_m$ for all k and m, but $c_k \neq c$, can be handled by estimating one parameter. The estimation is quite simple since what needs to be done is to scale all the d' values for the absolute-judgment task to the unit of measurement appropriate for the comparative-judgment task by multiplication with the parameter to be estimated.

In addition to the possibilities of nonindependence, unequal response bias in the comparative-judgment task, and differences in the unit of measurement for absolute- and comparative-judgment tasks, there is again the possibility that level of attention is not as high for each of n-alternatives as for one. Thus, the assumption that $t_{ik}(s') = s_i(s') - c_k$ is not as plausible as it might appear at first sight.

Furthermore, the maximum rule for comparative judgment can only be applied to the remembered values of t_{ik} for $k = 1,..., n$. If memory is not perfect, this adds an additional source of variance to the n-alternative comparative judgment that is not present in the absolute judgment. If the memory noise is dependent only on time and independent of the level of $s_k(s')$, it can be absorbed in the bias parameter, c_k, of course at the cost of insuring that $c_k \neq c_m$ for $k \neq m$. If the memory loss is a function of the level of $s_i(s')$, the $s_i(s')$ variables derived from the absolute-judgment task are no longer appropriate for predicting comparative judgments and must be modified to reflect the effects of memory loss.

EFFECTS OF NOISE IN THE INDEPENDENT VARIABLE ON THE DISTRIBUTION OF THE DEPENDENT VARIABLE

So far, we have defined s_i to be a random-variable function of a real variable s_i'. If we assume that the independent variable s_i' is a random variable, we are faced with the problem of defining what it means for a random variable to be a function of a random variable. Let us assume that we know the forms of the distribution functions of the two random variables. In this case, a random-variable function of a random variable means a random variable whose parameters are real-valued functions of the independent random variable.[2] Assuming that the distributions are normal, we obtain:

$$s_i \sim N[\mu_{s_i} = \mu_{s_i}(s_i' \sim N[\mu_{s_i'}, \sigma_{s_i'}]),$$

$$\sigma_{s_i} = \sigma_{s_i}(s_i' \sim N[\mu_{s_i'}, \sigma_{s_i'}])].$$

[2] The notion of a random variable being a function of another random variable is just a slightly different, but equivalent, way of talking about conditional probability distributions (see Parzen, 1962, pp. 41–65). Furthermore, the present section considers a very special case where the conditional distributions have the same (normal) form and variance, differing only in mean. When this special case is a reasonable approximation, the effects of noise in the independent variable are very easy to determine.

Solving for the distribution of s_i can in general be quite complicated, but there is one special case of considerable use where the solution is very simple. The assumptions are that, for virtually all of the values of s'_1 *within its range of variability*, the following functions are good approximations: $\mu_{s_i}(s'_i) = as'_i + b$ and $\sigma_{s_i}(s'_i) = \bar{\sigma}_i$, where a, b, and $\bar{\sigma}_i$ are real variables. Provided the noise in any given value (s'_i) of the independent variable is not "too" great, these approximations will be satisfactory, regardless of the nature of the functions $\mu_s(s')$ and $\sigma_s(s')$ over the entire possible range of s'.

The distribution of s_i in the above case is easily obtained by splitting s_i into the sum of two independent random variables x and y, where

$$x \sim N[\mu_{s_i}, 0] = N[a_i s'_i + b_i, 0],$$

and

$$y \sim N[0, \sigma_{s_i}] = N[0, \bar{\sigma}_i].$$

Now note that $x \sim N[a_i s'_i + b_i, 0]$ implies that $x = a_i s'_i + b_i$, which implies that $x \sim N[a_i \mu_{s'_i} + b_i, a\sigma_{s'_i}]$.

Thus

$$s_i = x + y \sim N[a_i \mu_{s'_i} + b_i, (a_i^2 \sigma^2_{s'_i} + \bar{\sigma}_i^2)^{1/2}],$$

since the sum of two independent normally distributed random variables is normally distributed with its mean and variance equal to the sum of the means and variances, respectively, of the components. Of course, it must be remembered that s_i, b_i, and σ_i are in general different for every value of s_i, and we have no fundamental interest in their values. What we care about is showing that the distribution of s_i is approximately normal. Note that the standard deviation of s_i, $(a_i^2 \sigma^2_{s'_i} + \bar{\sigma}_i^2)^{1/2}$, is a function of s'_i over the entire range of s'. Thus, we must choose the standard deviation of one particular s_i as the unit of measurement and measure every other s_j with this unit.

Incidentally, this same method of calculating the effects of noise in the independent variable on the distribution of a dependent variable which has noise in the mapping from independent to dependent variable, can be directly applied to two psychological (intervening) variables in a functional chain, $v_i = v_i(s_i)$. As long as the noise in the random variable s_i is not too large in relation to the rate at which μ_{v_i} is deviating from a linear function of s_i, and σ_{v_i} deviating from a constant function of s_i, the distribution of v_i will be normal, provided the distributions of the noise in s_i and the mapping from s_i to v_i are normal. This is quite convenient for strength theories with many levels of intervening variables organized in a functional chain.

CONCLUSION

Since strength theory uses the standard deviation of the total noise in some condition as the unit of measurement, it is necessary to consider the theoretical components of the total noise in order to be sure that the same unit of measurement is being applied to different conditions.

Whenever the judgment task and the criteria are certain to be identical for a number of conditions and one can plot every condition against the standard condition on an operating characteristic, the slopes provide one with the information as to how the noise is varying over the different conditions. In such cases, d' measures everything with the same unit of measurement, and the only function of a careful analysis of the components of variance is to provide a plausible and simple explanation of the differences in variance for different conditions.

However, when one attempts to make predictions from one judgment task to another, such as from "yes–no" to rating absolute judgment or from absolute to comparative judgment, where direct assessment of differences in variance from one task to another is not possible, it is extremely important that one have a plausible analysis of the components of variance in each case. With such an analysis it will generally be possible to determine what parameters will have to be estimated in order to make predictions from one task to another and it will be relatively clear how to estimate them. Also, it will be clear what data one must have to make an adequate test of the theory. Lacking such an analysis of the variance components, one is quite likely to reach erroneous conclusions.

REFERENCES

Atkinson, R. C., and Estes, W. K. Stimulus sampling theory. In R. D. Luce, R. R. Bush, and E. Galanter (Eds.), *Handbook of mathematical psychology*, Vol. 2. New York: Wiley, 1963. Pp. 121–268.

Atkinson, R. C., and Shiffrin, R. M. Mathematical models for memory and learning. Paper presented at the Third Conference on Learning, Remembering, and Forgetting, New York Academy of Sciences, Princeton, New Jersey, 1965.

Atkinson, R. C., Bower, G. H., and Crothers, E. J. *An introduction to mathematical learning theory*. New York: Wiley, 1965.

Bower, G. H. Notes on a descriptive theory of memory. Paper presented at the Second Conference on Learning, Remembering and Forgetting, New York Academy of Sciences, Princeton, New Jersey, 1964.

Bush, R. R., and Mosteller, F. *Stochastic models for learning*. New York: Wiley, 1955.

Creelman, C. D. Human discrimination of auditory duration. *Journal of the Acoustical Society of America*, 1962, **34**, 582–593.

Egan, J. P. Recognition memory and the operating characteristic. Indiana Univer., Hearing and Communication Lab., AFCRC-TN-58-51, AD-152650, 1958.

Egan, J. P., Schulman, A. I., and Greenberg, G. Z. Operating characteristics determined by binary decisions and by ratings. *Journal of the Acoustical Society of America*, 1959, **31**, 768–773.

Egan, J. P., Greenberg, G. Z., and Schulman, A. I. Interval of the time uncertainty in auditory detection. *Journal of the Acoustical Society of America*, 1961, **33**, 771–778.

Elliott, P. B. Tables of α'. In J. A. Swets (Ed.), *Signal detection and recognition by human observers*. New York: Wiley, 1964. Pp. 651–684.

Estes, W. K. Toward a statistical theory of learning. *Psychological Review*, 1950, **57**, 94–107.

Estes, W. K. Component and pattern models with Markovian interpretations. In R. R. Bush and W. K. Estes (Eds.), *Studies in mathematical learning theory*. Stanford: Stanford Univer. Press, 1959. Pp. 9–52.

Green, D. M. Psychoacoustics and detection theory. *Journal of the Acoustical Society of America*, 1960, **32**, 1189–1203.

Green, D. M., and Birdsall, T. G. The effect of vocabulary size on articulation score. (Univer. of Michigan: Electronic Defense Group, Technical Memorandum, No. 81, 1958). In J. A. Swets (Ed.), *Signal detection and recognition by human observers*. New York: Wiley, 1964.

Green, D. M., and Swets, J. A. *Signal detection theory and psychophysics*. New York: Wiley, 1966.

Hull, C. L. *Principles of behavior*. New York: Appleton, 1943.

Hull, C. L., Hovland, C. I., Ross, R. T., Hall, M., Perkins, D. T., and Fitch, F. G. *Mathematico-deductive theory of rote learning*. New Haven, Connecticut: Yale Univer. Press, 1940.

Luce, R. D. *Individual choice behavior*. New York: Wiley, 1959.

Luce, R. D. A threshold theory for simple detection experiments. *Psychological Review*, 1963, **70**, 61-79. (a)

Luce, R. D. Detection and recognition. In R. D. Luce, R. R. Bush, and E. Galanter (Eds.), *Handbook of mathematical psychology*, Vol. 1. New York: Wiley, 1963. Pp. 103-189. (b)

Markowitz, J., and Swets, J. A. Factors affecting the slope of empirical ROC curves: comparison of binary and rating responses. *Perception and Psychophysics*, 1967, **2**, 91–100.

Nachmias, J., and Steinman, R. M. Study of absolute visual detection by the rating-scale method. *Journal of the Optical Society of America*, 1963, **53**, 1206–1213.

Norman, D. A. Sensory thresholds, response biases, and the neural quantum theory. *Journal of Mathematical Psychology*, 1964, **1**, 88–120.

Norman, D. A., and Wickelgren, W. A. Short-term recognition memory for single digits and pairs of digits. *Journal of Experimental Psychology*, 1965, **70**, 479–489.

Parzen, E. *Stochastic processes*. San Francisco: Holden-Day, 1962. Pp. 41–65.

Pollack, I. Identification of elementary auditory displays and the method of recognition memory. *Journal of the Acoustical Society of America*, 1959, **31**, 1126–1128.

Schulman, A. I., and Greenberg, G. Z. Operating characteristics and *a priori* probability of the signal. *Journal of the Acoustical Society of America*, 1960, **32**, 1505(A).

Schulman, A. I., and Mitchell, Rhoda R. Operating characteristics from yes-no and forced-choice procedures. *Journal of the Acoustical Society of America*, 1966, **40**, 473–477.

Shipley, E. F. Detection and recognition: experiments and choice models. *Journal of Mathematical Psychology*, 1965, **2**, 277–311.

Swets, J. A. Indices of signal detectability obtained with various psychophysical procedures. *Journal of the Acoustical Society of America*, 1959, **31**, 511–513.

Swets, J. A., and Birdsall, T. G. The human use of information. III. Decision making in signal detection and recognition situations involving multiple alternatives. *IRE* (*Institute of Radio Engineers*), *Transactions on Information Theory*, 1956, IT-2, 138–165.

SWETS, J. A., TANNER, W. P., JR., AND BIRDSALL, T. G. The evidence for a decision-making theory of visual detection. Univer. of Michigan, Electronic Defence Group, Technical Report No. 40, 1955.

SWETS, J. A., TANNER, W. P., JR., AND BIRDSALL, T. G. Decision process in perception. *Psychological Review*, 1961, **68**, 301–340.

TANNER, W. P., JR. Theory of recognition. *Journal of the Acoustical Society of America*, 1956, **28**, 882–888.

TANNER, W. P., JR., AND NORMAN, R. The human use of information. II. Signal detection for the case of an unknown signal parameter. *IRE (Institute of Radio Engineers), Transactions on Professional Group on Information Theory*, 1954, PGIT-4, 222–227.

TANNER, W. P., JR., AND SWETS, J. A. A decision-making theory of visual detection. *Psychological Review*, 1954, **61**, 401–409.

THURSTONE, L. L. Law of comparative judgment. *Psychological Review*, 1927, **34**, 273–286.

WATSON, C. S., RILLING, M. E., AND BOURBON, W. T. Receiver-operating characteristics determined by a mechanical analog to the rating scale. *Journal of the Acoustical Society of America*, 1964, **36**, 283–288.

WEINTRAUB, D. J., AND HAKE, H. W. Visual discrimination, an interpretation in terms of detectability theory. *Journal of the Optical Society of America*, 1962, **52**, 1179–1184.

WICKELGREN, W. A. Consolidation and retroactive interference in short-term recognition memory for pitch. *Journal of Experimental Psychology*, 1966, **72**, 250–259.

WICKELGREN, W. A., AND NORMAN, D. A. Strength models and serial position in short-term recognition memory. *Journal of Mathematical Psychology*, 1966, **3**, 316–347.

RECEIVED: September 6, 1966