Sequential dependencies and regression in psychophysical judgments*

DAVID V. CROSS

State University of New York at Stony Brook, Stony Brook, New York 11790

A tendency for judgments of stimulus magnitude to be biased in the direction of the value of the immediately preceding stimulus is found in magnitude estimations of loudness. This produces a bias in the empirical psychophysical function that results in underestimation of the exponent of the unbiased function presumed to relate number and stimulus intensity, $N = aS^n$. The biased judgment can be represented as a power product of focal and preceding stimulus intensity, $N_{ij} = aS_i^m S_j^p$. A bias-free estimate of the correct exponent, n, can be obtained from the relation $n = m + b$.

Recently, Stevens and Greenbaum (1966) called attention to the possible nonequivalence of magnitude estimation and production functions. In cross-modality matching experiments in general, different power function relationships are usually obtained, depending on which variable the S controls. Ss appear to shorten the range of their adjustments on whichever variable is under their control, thereby producing two regression lines (in log-log coordinates) and two exponents for the power function presumed to relate the two variables. This has been called a "regression effect" because of its similarity to the fact that two regression lines are obtained in bivariate distributions whenever the correlation between two variables is less than perfect, and also because of its similarity to the problem of estimating regression parameters when measurements on both variables are in error; two regression lines are typically obtained in the latter case, depending on whether the vertical or horizontal error sum of squares is minimized in fitting the best line, and both lines are biased (Wald, 1940).

Neither the bivariate model, with statistical regression toward the mean, nor the errors in both variables linear regression model, with a least squares estimator that underestimates the true regression parameter, seem to be appropriate models for this apparently ubiquitous effect in cross-modality matching experiments. The independent variables in such experiments are not allowed to vary freely. They are under the E's control and are presented at predetermined levels. Even when presented with error, levels of the independent variable can be legitimately considered as fixed in the ensuing analysis. The regression model appropriate in these instances is one proposed by Berkson (1950). Not two, but one regression line is obtained by the method of least squares, and it is an unbiased estimate of the true line.

Consequently, the so-called "regression effect" in psychophysical judgments arises not from the imperfect dependence of one variable on another; nor does it arise

from errors in measuring stimulus levels, presented or produced, or from the inevitable noisiness of the matching judgments themselves. Such effects, when they exist, must be due to the occurrence of systematic biases in human judgments.

Hollingworth (1910) called attention to one such bias, a "centering tendency," the tendency for judgments to gravitate toward a mean magnitude, thereby shortening the range of the dependent variable. Garner (1953) reported findings suggesting that this centering tendency is traceable to sequential dependencies in judgments. Ss required to categorize up to 20 loudness levels showed a tendency to give a higher rating to a stimulus when the preceding stimulus was higher than that judged, and a lower rating when the preceding stimulus was lower than that judged. This tendency toward "assimilation" produces an overall "centering effect" because of the unavoidable fact that, in a given stimulus series, higher stimuli must, on the average, be preceded by lower stimuli, and lower stimuli by higher stimuli.

Assimilative tendencies in absolute judgments and in category judgments have been demonstrated by Holland and Lockhead (1968), Ward and Lockhead (1970, 1971), and by Ward (1972). In addition, Ward (1973) showed that these tendencies are not peculiar to absolute category judgments, but that they can also occur in magnitude estimation data.

The present study shows that assimilative sequential dependencies in magnitude judgments can contribute to the regression effect through a systematic bias in judgment that results in underestimation of the true power law exponent.

METHOD

Eighteen unpaid volunteers, Harvard graduate and undergraduate students, served as Ss in single sessions lasting approximately 10 min. All had previous experience in psychophysical experiments but were not told the purpose of this one. They were instructed simply to assign numbers proportional to the apparent loudness of presented noise bursts. No standard stimulus or modulus was designated. Ss were seated in an I.A.C. sound-attenuating chamber. Judgments were given orally into a boom microphone mounted on S's headset and were monitored by E.
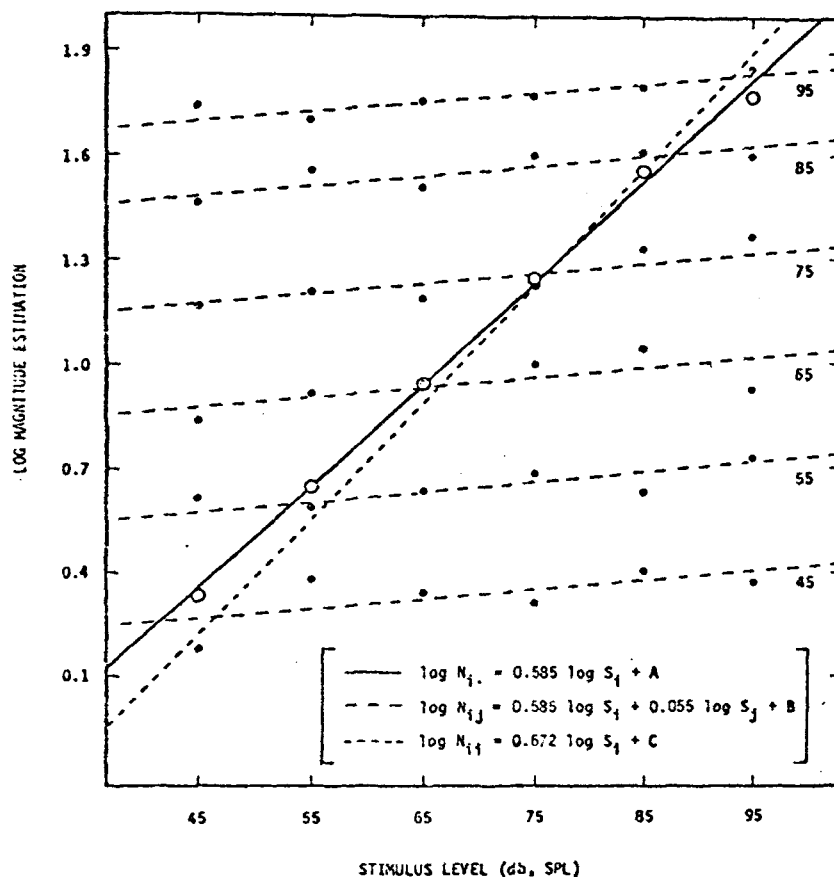
Fig. 1. Apparent loudness of standard noise as a function of immediate stimulus level, S_i, and of the level of the immediately preceding stimulus, S_j. See text for explanation of equations.

The stimuli were 1-sec bursts of standard noise (USASI spectrum; from a General Radio noise generator, Model 1382) separated uniformly by 7.5-sec intervals. They were gated with a 15-msec rise-decay time and presented binaurally (Koss ESP-9 earphones) at six intensity levels (45 to 95 dB SPL in 5-dB steps) in six quasirandomly permuted orders, plus an extra stimulus in initial position, for a total of 37 stimulus presentations. The schedules were so designed that each stimulus was preceded exactly once by every other stimulus, including itself. Following a Latin square design, each S received a different presentation schedule, so that second- and higher-order sequential effects could be counterbalanced and so that every stimulus appeared in each initial position and in every other position an equal number of times.

RESULTS

Power Function Fit

It is recommended practice in the analysis of magnitude estimation data to use geometric means in fitting the psychophysical function (Stevens, 1971). One reason for this is that if individual Ss obey power functions, then this averaging procedure determines a group function that will be of the same form with an exponent that is the arithmetic average of the individual exponents. Also, both the dispersion and skewness of judgments tend to increase in direct proportion to average magnitude, a characteristic of log normal distributions. The geometric mean corresponds to the

median log normal variate, and its logarithm is a minimum-variance unbiased estimator of the distribution's location parameter. The sample median is unbiased, but it is less efficient. The arithmetic mean, on the other hand, is biased.

Geometric means of the 108 magnitude estimations given at every stimulus level (18 Ss x 6 replications of each level) are represented in Fig. 1 by open circles. The straight line drawn through these points was fitted conventionally by the method of least squares, and it has a slope equal to 0.585. With a standard error of 0.015, this estimate of the exponent for the power function relating loudness to sound pressure is significantly smaller $(p < .005)$ than the value 0.67 currently proposed as typical (Stevens, 1972).

Sequential Effects

A more detailed examination of the data reveals the influence of contextual or preceding stimulus levels on magnitude estimation of focal stimuli. The small filled circles in Fig. 1 show judged magnitude as a function of the preceding stimulus level. Each symbol represents the geometric mean of 18 independent judgments and corresponds to N_{ij}, the number (averaged over Ss) assigned to Loudness Level i, given the immediately preceding stimulus of Loudness Level j (i,j = 45, 55, 65,

75, 85, 95 dB SPL). The curves approximated by dashed lines passing through these points represent how estimates tend to increase (or decrease) as the preceding stimulus increases (or decreases). These lines have a common slope equal to 0.055. A comparison of the fit obtained when each least squares regression line is allowed a different slope with that obtained when all are constrained to have a common slope equal to the average regression coefficient constitutes a test for parallelism (Sprent, 1969, p. 100). The results of this test support the hypothesis that individual slopes are equal, $F(5,24) = 0.362$. The important effect summarized by these lines is that judgments of stimulus magnitude are assimilated toward the value of the immediately preceding stimulus.

This effect is small when measured in terms of variance components. The proportion of variance in log judgment accounted for by linear regression on log stimulus intensity is 0.982. When the second variable, log contextual stimulus intensity, is added to the regression equation, the proportion of variance in the dependent variable accounted for by multiple regression on the two independent variables is increased to 0.991. This means that taking this second orthogonal variable into account adds less than 1% to the total proportion of response variance explained by the best fitting power function.

However small this order effect, its net result is underestimation of the theoretically correct power function exponent. In Fig. 1, for example, each large open circle represents the average of the corresponding small circles. These averages are biased for the reasons mentioned above: The bias is upward for low stimuli and downward for high stimuli. The net effect on the regression line relating average log judged magnitude to log physical magnitude is a reduction in its slope relative to that of a line through points free of order bias. The greater the order bias, the greater the reduction in slope. But it is the slope of precisely this line that is customarily taken as the "best" estimate of the exponent in the power function relating sensory magnitude to physical magnitude. Consequently, exponents are underestimated when order effects are present.

A Descriptive Model

How we compensate for this order bias in the search for the presumably correct power function depends on how the problem is modeled. Presumably, the model should treat magnitude estimations of stimuli preceded only by themselves as free of bias. Also, since the dashed lines in Fig. 1 approximating the apparent trend with respect to preceding stimulus level are all parallel, the bias should be a monotonic function of the size of the ratio of the preceding stimulus to the focal stimulus.

The simplest model incorporating these constraints is represented by Eq. 1 below. In terms of the focal stimulus, S_i, and the immediately preceding stimulus, S_j, it asserts that numerical judgment, N_{ij}, is proportional to the true power function relating sensory magnitude to stimulus intensity, multiplied by a bias factor which is proportional to the ratio S_j/S_i raised to a power, b, the sign and size of which determines the direction and magnitude of the effective bias. This is written as follows:

$$N_{ij} = aS_i^n(S_j/S_i)^b. \qquad (1)$$

For positive b, this model produces assimilation, and for negative b, it results in contrast. For a given b, the magnitude of the bias is monotonically increasing with the separation between the focal and contextual stimulus levels; when the levels are identical, the judgment is bias free. The exponent, n, represents the true exponent, not that value which would be obtained by fitting via conventional procedures a function to the geometric means, N_i, represented in Fig. 1 by open symbols. Each of these six data points represents a geometric mean of six values of N_{ij}, one value for each preceding stimulus level. Thus

$$N_i = (\prod_j N_{ij})^{1/6}. \qquad (2)$$

By rewriting Eq. 1 and substituting in Eq. 2, we obtain

$$N_i = aS_i^{n-b}(\prod S_j^b)^{1/6}. \qquad (3)$$

Since, as in the above presentation schedule, the factor in parenthesis can be made constant for all i regardless of b, it can be absorbed by the proportionality constant and we can write

$$N_i = kS_i^m, \qquad (4)$$

where $m = n - b$ and $k = a(\prod S_j^b)^{1/6}$. Equation 4 expresses the function conventionally fitted to magnitude estimation data. But the exponent, m, underestimates the theoretically correct exponent, n, by the size of the bias component, b. Additional variance about the regression curve (and even systematic contribution to bias) may arise by failure to equate the stimulus contexts for each level; that is, k may vary with S_i, depending on the context, and this dependence may be systematic.

A means of analysis is called for that will provide estimates of both n, the desired exponent, and b, the bias parameter. Equation 1 can be rewritten as a power product involving both focal and preceding stimuli,

$$N_{ij} = aS_i^m S_j^b, \qquad (5)$$

remembering that $m = n - b$. For the data depicted in Fig. 1, this is a simple three-variable regression problem in which least squares estimates of m and b can be obtained by fitting the following linear regression equation:
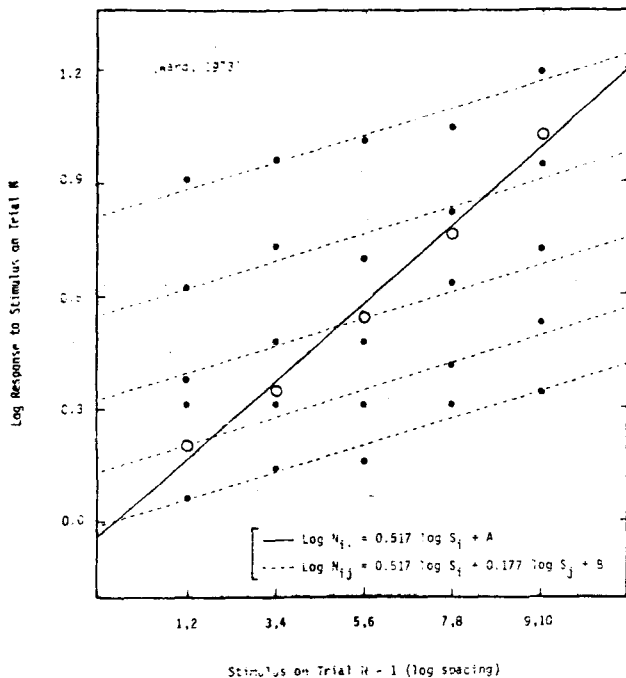
Fig. 2. Sequential effects in magnitude estimations of the loudness of 1,000-Hz tones. Data from Ward (1973).

$$\log N_{ij} = \log a + m \log S_i + b \log S_j. \qquad (6)$$

This was done for the data at hand, with the following results:

$$m = 0.585 \quad \text{and} \quad b = 0.055.$$

The first regression coefficient,[*] m, represents the slope of the least squares line drawn through the open symbols in Fig. 1. The dashed lines passing through the filled symbols all have the common slope, b = 0.055. An adjusted estimate of the loudness exponent is n = m + b = 0.64, a value that is not significantly different from the proposed theoretical value of 0.67.

An alternative way of obtaining a bias-free estimate of this exponent is to discard the biased responses and simply fit the appropriate function to an unbiased subset of the data; namely, to those responses evoked by stimuli preceded by themselves in the presentation schedule. This is highly wasteful of data, but the model regards only these responses as bias free. Using Eq. 5 and setting i = j, we obtain

$$N_{ii} = aS_i^{n-b} S_i^b = aS_i^n. \qquad (7)$$

The finely dashed line running diagonally through Fig. 1 represents the least squares fit to the six data points entering the above equation. The least squares estimate of n in this case turns out to be 0.672; this estimate has a standard error of 0.023.

DISCUSSION

Since this initial experiment was undertaken, the basic paradigm has been replicated several times, both in classroom demonstrations and in controlled laboratory scaling experiments. In all demonstrations and experiments in which an atypically low exponent was obtained, the data also exhibited the systematic assimilative order effects exemplified by the present study. The amount of bias may differ from one experiment to another and across modalities (the bias has also been observed in sequential judgments of brightness, vibration on the fingertip, and of lifted weights), but the direction appears to be consistently that of the assimilation of response magnitude toward a previous value.

The largest sequential bias effect thus far detected appears in the results obtained by Ward (1973) from magnitude estimations of the loudness of 1,000-Hz tones. Quantitative assessment of the bias was not attempted by Ward, but, by applying the present model to stimulus and response values read off the graphs published by him, it is possible to repeat the analysis performed here and to see how well the model of Eq. 1 can handle these results.

The least squares estimate of the loudness exponent for the data depicted in Fig. 1 of the cited study is equal to 0.51 and has a standard error equal to 0.016. This estimate is unusually low, but examination of the sequential properties of the data provides an explanation of why such a small value was obtained.

The sequential effects observed by Ward are exhibited in Fig. 2. This graph is a redrawing of Ward's Fig. 2 to show the results in log-log coordinates. A test for homogeneity of the individual bias regression coefficients supports the hypothesis of parallelism; $F(4,17) = 0.667$. Consequently, the dashed lines showing the trend of this bias have all been drawn with a common slope corresponding to the average regression coefficient, which represents the bias parameter, b, of Eq. 1; it is equal to 0.177. This value is three times the size of the effect obtained in the present study.

It may be pointless to speculate on the procedural differences in the two studies that might have contributed to this large difference. Ward notes that his method of instructing S to judge the ratio between the present and previous stimulus and to assign a number that bears the same ratio to the number previously assigned should be particularly subject to sequential bias.

A revised estimate of the loudness exponent is obtained by adding the two empirical regression constants, m and b of Eq. 5. The result is 0.510 + 0.177 = 0.687, a value not significantly different from the proposed theoretical value of 0.67.

Ward presents a summary (his Table 2) of average log response ratios for all ordered pairs of stimulus values occurring successively in his experiment. These measures

provide an opportunity to examine other implications of the bias model. For example, the diagonal elements of this table are all logarithms of the ratios N_{ii}/N_i—the geometric mean response to S_i, when preceded by itself, divided by the geometric mean response to S_i, regardless of previous stimulus. Using Eqs. 7 and 4, the following identity can be written:

$$N_{ii}/N_i = a(S_i^{n-b}S_i^b)/S_i^{n-b} = aS_i^b. \qquad (8)$$

Thus, the diagonal elements should be proportional to the product of the bias parameter and the log of the focal stimulus value; that is, $\log N_{ii}/N_i = b \log S_i + \log a$. This relationship for Ward's data is plotted here in Fig. 3. The least squares line in this graph has a slope equal to 0.179—a value in close agreement with that of the above analysis.

A similar analysis can be extended to all pairs of adjacent response ratios. It turns out, however, to be simpler to take the geometric means of complementary pairs, i.e., $[(N_{ij}/N_j)(N_{ji}/N_i)]^{1/2}$. For these measures, the order effects model implies an engagingly simple relationship:

$$[(N_{ij}/N_j)(N_{ji}/N_i)]^{1/2}$$

$$= a[(S_i^{n-b}S_j^b)(S_j^{n-b}S_i^b)/(S_j^{n-b})(S_i^{n-b})]^{1/2}$$

$$= a(S_iS_j)^{b/2}. \qquad (9)$$

For the 55 unordered stimulus pairs involved, a correlation of 0.962 is obtained between the log of the geometric mean of the response ratios for complementary pairs of stimuli, Ward's "symmetry numbers," and $b(\log S_i + \log S_j)/2$, the linear relationship in log coordinates implied by Eq. 9. This relationship explains why Ward obtained a high correlation $(r = 0.947)$ between the "symmetry numbers" and total strength of the two stimuli (the sum of the stimulus magnitudes in millivolts). The linear regression of the "symmetry numbers" against $(\log S_i + \log S_j)/2$ yields 0.19 as an estimate of the bias parameter, b. This value is not significantly different from the values 0.179 or 0.177 previously calculated. The invariances indicated in these analyses are all the more impressive when it is recognized that the mean responses, N_i (or N_j) entering the above equations are not unbiased estimators of S_i^{n-b} as asserted in Eq. 4, but, rather, are subject to second-order sequential dependencies arising from the fact that preceding stimuli, themselves, were not all preceded by every other stimulus.

The existence of assimilative effects in psychophysical experiments may be more widespread than indicated by the relatively small number of experiments designed explicitly to reveal them. The same effects in serial integration tasks are called recency effects (Anderson, 1971). Anderson reports that when a S lifts three
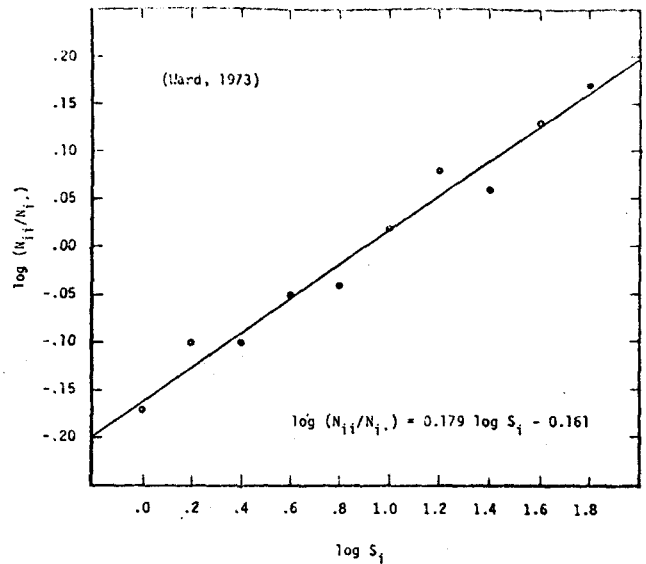

Fig. 3. Measurement of order effects bias from an experiment by Ward (1973). Each ordinate represents the log of the geometric mean of the response ratios N_{ii}/N_{ij}, where N_{ii} is the average loudness of a tone, S_i, when preceded by itself and N_{ij} is the average loudness of a tone preceded by an arbitrary level, S_j.

weights in sequence under instructions to judge the heaviness of the last weight, the judgments show a positive context (assimilation) effect.

When assimilation or recency effects occur in bisection experiments, they are identified as hysteresis effects (Stevens, 1957). In the classic bisection paradigm, three stimuli, say, ⟨a, x, b⟩, are presented one at a time in either increasing or decreasing magnitude. The first and third stimulus, a and b, are predetermined, but the middle stimulus, x, is variable and is adjusted to be halfway between a and b. If the apparent magnitude of x is assimilated toward the value of the preceding stimulus, a, it must be adjusted away from a, toward b, in order to appear subjectively halfway between them. When the sequence is reversed, with b preceding x, the opposite occurs. This is what typically happens in bisection.

In some experiments, Ss have been asked to give magnitude estimations of the difference between a pair of stimuli (Beck & Shaw, 1967; Curtis, Attneave, & Harrington, 1968; Fagot & Stewart, 1969; Dawson, 1971). Sensory scales derived from these studies behave like partition scales in that they are nonlinear relative to scales created by matching numbers directly to single stimuli. The scales are usually power functions of the corresponding physical variable, but the exponents are much smaller than expected; they compare in value with exponents obtained from partition procedures such as category judgments and equisection, or with exponents obtained from experiments in which a large regression effect was observed. Stevens (1971) contends that scales for prothetic continua derived from interval estimation procedures are of necessity biased, since a constant

subjective interval must, as the defining stimulus levels are raised, eventually fall below the limits of the sensory system's resolving power. Consequently, a constant sensory difference seems smaller and smaller as it is moved up the continuum, until, finally, it becomes smaller than a jnd and simply disappears. To compensate for this, Os judge intervals *as if* the underlying power function had a virtual exponent about half as large as the actual exponent of the continuum. This is a simple descriptive account of the findings, and it represents one way to model the bias that is presumed to be operating in this paradigm. An alternative bias model might be derived from the assumption of assimilative effects operating in a symmetric fashion between a pair of stimuli presented either repeatedly in succession or presented simultaneously. The kind of effects that were observed in this study would-make stimuli appear less different than the simple subtraction of appropriate power function scale values would predict. The simple algebraic difference model may have to be abandoned, where prothetic continua are involved, for a somewhat more complicated model involving an interaction between the members of a stimulus pair. Indeed, the evidence based on tests of axioms of additivity and of multiplicativity seem to point in this direction (Fagot & Stewart, 1969). The problem requires further investigation.

It would be nice to be able to conclude that the major contribution of this study is the solution of the regression problem in psychophysics. Indeed, it does show that order bias is a sufficient condition for psychophysical regression, and such a bias may well be a major component of regression whenever regression occurs, but it has not shown that order bias is a necessary condition for that effect. For example, unpracticed Os may simply feel constrained to avoid extreme numbers in their judgments. That, too, may be a sufficient condition for regression, independent of order effects.

The question that remains to be answered is whether or not other forms of cross-modality matching—including the method of magnitude production—are also biased by order effects. There is reason to believe that in most magnitude production procedures order effects would not be a problem. In adjusting loudness to match number, for example, it may be necessary to repeat approximations to a match a number of times before a final match is found. These conditions may cancel a potential order bias, just as in the present study judgments of repeated stimuli tend to be free of bias.

REFERENCES

Anderson, N. H. Test of adaptation-level theory as an explanation of a recency effect in psychophysical integration. Journal of Experimental Psychology, 1971, 87, 57-63.

Beck, J., & Shaw, W. A. Ratio-estimations of loudness-intervals. American Journal of Psychology, 1967, 80, 59-65.

Berkson, J. Are there two regressions? Journal of the American Statistical Association, 1950, 45, 164-180.

Curtis, D. W., Attneave, F., & Harrington, T. L. A test of a two-stage model of magnitude judgment. Perception & Psychophysics, 1968, 3, 25-31.

Dawson, W. E. Magnitude estimation of apparent sums and differences. Perception & Psychophysics, 1971, 9, 368-374.

Fagot, R. F., & Stewart, M. Tests of product and additive scaling axioms. Perception & Psychophysics, 1969, 5, 117-123.

Garner, W. R. An informational analysis of absolute judgments of loudness. Journal of Experimental Psychology, 1953, 46, 373-380.

Holland, M. K., & Lockhead, G. R. Sequential effects in absolute judgments of loudness. Perception & Psychophysics, 1968, 3, 409-414.

Hollingworth, H. L. The central tendency of judgment. Journal of Philosophy & Psychology, 1910, 7, 461-469.

Sprent, P. *Models in regression and related topics.* London: Methuen, 1969.

Stevens, S. S. On the psychophysical law. Psychological Review, 1957, 64, 153-181.

Stevens, S. S. On predicting exponents for cross-modality matches. Perception & Psychophysics, 1969, 6, 251-256.

Stevens, S. S. Issues in psychophysical measurement. Psychological Review, 1971, 78, 426-450.

Stevens, S. S. Perceived level of noise by Mark VII and decibels (E). Journal of the Acoustical Society of America, 1972, 51, 575-601.

Stevens, S. S., & Greenbaum, H. B. Regression effect in psychophysical judgment. Perception & Psychophysics, 1966, 1, 439-446.

Wald, A. The fitting of straight lines if both variables are subject to error. Annals of Mathematical Statistics, 1940, 11, 284-300.

Ward, L. M. Category judgments of loudnesses in the absence of an experimenter-induced identification function: Sequential effects and power function fit. Journal of Experimental Psychology, 1972, 94, 179-184.

Ward, L. M. Repeated magnitude estimations with a variable standard: Sequential effects and other properties. Perception & Psychophysics, 1973, 13, 193-200.

Ward, L. M., & Lockhead, G. R. Sequential effects and memory in category judgments. Journal of Experimental Psychology, 1970, 84, 27-34.

Ward, L. M., & Lockhead, G. K. Response system processes in absolute judgment. Perception & Psychophysics, 1971, 9, 73-78.