Journal of Experimental Psychology: General
1980, Vol. 109, No. 3, 304-319

Comparison of Two Theories of "Ratio" and "Difference" Judgments

Michael H. Birnbaum
University of Illinois at Urbana-Champaign

This article examines the hypothesis that judges compare stimuli by ratio and subtractive operations when instructed to judge "ratios" and "differences." Rule and Curtis hold that magnitude estimations are a power function of subjective values, with an exponent between 1.1 and 2.1. Accordingly, the two-operation model tested assumes magnitude estimations of "ratios" are a comparable power function of subjective ratios. In contrast, Birnbaum and Veit theorize that judges compare two stimuli by subtraction for both "ratio" and "difference" instructions and that magnitude estimations of "ratios" are approximately an exponential function of subjective differences. Three tests were used to compare the theory of one operation with the two-operation theory for the data of nine experiments. The results strongly favor the theory that observers use the same operation for both instructions.

The most puzzling and challenging problem in psychophysics, and perhaps the biggest headache, has been the failure of convergent operations to define a single scale of sensation. Scaling methods based on "ratio" instructions and ratio models usually yield scales that differ drastically from the scales obtained with "interval" instructions and subtractive models.[1] For example, magnitude estimations and category ratings of the same stimuli are often nonlinearly related.

Several theories have been advanced to account for this discrepancy. For example, Stevens (1971) argued that humans are unable to compute differences and that category ratings are "biased." Attneave (1962) hypothesized that numerical magnitude estimations are nonlinearly related to subjective values because the subjective values of numbers are nonlinearly related to objective number. Torgerson (1961) theorized that observers compare two stimuli in the same way regardless of "difference" or "ratio" instructions.

These theories could not be strongly distinguished in the early research on this problem. For example, Torgerson's (1961) theory that judges perform the same operation for "ratios" and "differences" cannot be tested on ordinal grounds without a multifactorial experiment. In a unifactor experiment, actual ratios and differences are monotonically related, since X/Y and $X - Y$ are monotonically related when Y is a constant. However, if X and Y are independently (e.g., factorially) manipulated, X/Y and $X - Y$ will not be monotonically related. For example, $4/1 > 7/3$ but $4 - 1 < 7 - 3$. Thus, if observers do perceive two relations between a pair of stimuli, judgments of "ratios" and "differences" should *not* be monotonically related in general, but should show the appropriate differences in rank order (Birnbaum & Veit, 1974a; Krantz, Luce, Suppes, & Tversky, 1971).

A number of recent studies using factorial

Thanks are due Barbara Mellers, Bernadette Setiadi, Stanley J. Rule, and Clairice T. Veit for helpful suggestions on the manuscript, and to Michael Levine for useful discussions.

Requests for reprints should be addressed to Michael H. Birnbaum, Department of Psychology, University of Illinois, Champaign, Illinois 61820.

[1] Quotation marks are used to denote instructions to the observer to judge "ratios" or "differences" and for numbers obtained using these instructions. Quotes are not used for theoretical statements, models, or for actual (numerical) ratios and differences.

designs conclude that judgments of "ratios" and "differences" are monotonically related, consistent with Torgerson's (1961) suggestion that observers instructed to judge "ratios" and "differences" compare the stimuli in the same way for both tasks (Birnbaum, 1978; Birnbaum & Elmasian, 1977; Birnbaum & Mellers, 1978; Birnbaum & Veit, 1974a; Hagerty & Birnbaum, 1978; Rose & Birnbaum, 1975; Schneider, Parker, Farrell, & Kanow, 1976; Veit, 1978; Elmasian & Birnbaum, Note 1).

In a comment on Veit's (1978) paper, Rule and Curtis (1980) argued that judges may use two operations but that perhaps the experiments have failed to detect them. They analyzed simulated data that were generated by two operations in an attempt to show how one might reach the erroneous conclusion that only one operation was present. Veit (1980) replied that the methods of data analysis used by Rule and Curtis were not applied by Veit (1978), that the procedure used by Veit (1978) correctly diagnoses their simulated data, and that their simulated data do not adequately reproduce Veit's (1978) empirical data.

Although they have not proposed a theory of "ratio" judgments and have not published tests of ratio models, Rule and Curtis have proposed a two-stage model that relates magnitude estimations of "differences" to magnitude estimations of single stimuli. The present article addresses the following question: Can the Rule and Curtis theory of magnitude estimation be successfully extended to account for judgments of "ratios" and "differences" with the two corresponding operations?

Analyses designed to highlight the difference between the theory of Birnbaum and Veit (1974a) and the theory of two operations are presented in the present article. Data for nine experiments are reviewed, using metric and nonmetric analyses that have greater power for distinguishing theories than the procedure of Rule and Curtis (1980). It will be shown that the predictions of the two theories are quite distinct and, furthermore, that the data strongly favor the hypothesis that judges use only one operation for both tasks.

Null Hypothesis: No Effect of Instructions

The following analogy may facilitate discussion. In the classical analysis of an experiment, the experimenter manipulates an independent variable and looks for a corresponding change, or difference, in the dependent variable. When a substantial effect is observed, one that would be improbable given the null hypothesis, the null hypothesis can be rejected in favor of the rival hypothesis that the independent variable has an effect on the dependent variable. However, when a difference is not observed, both the null hypothesis and the alternate are retained. No experiment can conceivably refute the general alternate hypothesis that a very tiny effect remains as yet undetected. However, as the power of the experiment grows, the magnitude of the likely effect is restricted by the data. Furthermore, one can reject specific alternatives, for example, that the effect exceeds some small value. From a Bayesian viewpoint, the plausibility of the null hypothesis grows with repeated experiments finding negligibly small effects.

In this analogy, the *independent variable* corresponds to the differential instructions to judge "ratios" or "differences", the *dependent variable* is the rank order in the factorial matrix, the *null hypothesis* is that instructions have no effect on the rank order, and the *alternative* is that the instructions will produce rank orders compatible with ratio and difference models. The *power* of the experiment depends on the ability to distinguish ratio from subtractive orders.

Birnbaum and Veit are in the uncomfortable position of defending the null hypothesis. But sometimes, as in the Michaelson-Morley experiment, the finding of no difference is the surprising one (Einstein, 1961). Michaelson and Morley were unable to detect a measurable difference in the speed of light dependent on the direction of the earth's motion. Their experiments could not disprove the hypothesis that a very small difference does exist. However, their experiments were a severe blow to the classical theory of mechanics based on the assumption that the earth moves through a fixed ether—a theory that predicted a sizable effect. Fortunately, the theory of Attneave, Rule, and Curtis can be extended to provide a specific alternative with which the Birnbaum-Veit theory can be compared and allows one to discuss the power of the experiments to discriminate between the two theories.

Two Theories of "Ratios" and "Differences"

Attneave (1962) theorized that magnitude estimation represents a special case of cross-modality matching, in which the subjective values of numbers are matched to the subjective values of the stimuli scaled. Attneave postulated that the observer in magnitude estimation selects a number, ϕ_N, such that the subjective value of the number, s_N, is equal to the subjective value, s_C, of the stimulus being scaled, ϕ_C. If $s_N = H_N(\phi_N)$ is the psychophysical function for number and $s_C = H_C(\phi_C)$ is the psychophysical function for the continuum being scaled, and if $s_N = s_C$, then $H_N(\phi_N) = H_C(\phi_C)$, or $\phi_N = H_N^{-1}\ [H_C(\phi_C)]$; thus, $\phi_N = H_N^{-1}(s_C)$. Therefore, the output (judgment) function for magnitude estimation, which relates subjective value to overt numerical response, should be the inverse of the input (psychophysical) function for number.

If the psychophysical function for number, H_N, is a power function ($s_N = \phi_N{}^n$), Attneave noted, the exponents obtained in magnitude estimation studies could be "off" by a multiplicative factor ($\phi_N = \phi_C{}^{c/n}$) and their ratios would still predict exponents from cross-modality matching experiments. Attneave proposed that category ratings produce equal intervals, though some have questioned why one kind of numerical judgment (category ratings) should differ from another (magnitude estimation). Based on limited evidence, Attneave estimated the exponent for numerals to be about .4. Subsequent developments by Curtis, Attneave, and Harrington (1968) and others (summarized by Rule and Curtis, in press) revised this estimate upward but retained the general theory.

Attneave's theory can be extended as follows. If the psychophysical functions are assumed to be power functions, $s = \phi^k$, where k is the exponent for the psychophysical function, and if judges use two operations for "ratio" and "difference" instructions, then magnitude estimations of "ratios" (MR) and magnitude estimations of "differences" (MD) might be represented as follows:

$$\mathrm{MR}_{ij} = a_R(\phi_j{}^k/\phi_i{}^k)^m + b_R, \qquad (1)$$

and

$$\mathrm{MD}_{ij} = a_D(\phi_j{}^k - \phi_i{}^k)^m + b_D, \qquad (2)$$

where $m = 1/n$ (the output function is the inverse of the psychophysical function for numerals). The constants a_R and a_D reflect such factors as the modulus and the standard "difference." The constants b_R and b_D allow for order effects or "response bias" in scale usage (e.g., if the subject compares a stimulus with itself, the subject might not judge the "ratio" to be 1 or the "difference" to be 0).[2]

Research of Curtis and Rule

In the research of Curtis and Rule with magnitude estimations, the standard stimulus, ϕ_i, is a constant or unspecified. Equation 1 is then written:

$$\mathrm{ME}_j = a\ \phi_j{}^{km} + b, \qquad (3)$$

where ME_j is the magnitude estimation of stimulus j, and a and b are constants. If the input transformation (represented by the exponent k) is the same for both tasks, and if the output exponent for magnitude estimation (m) is the same for both tasks, then the exponent estimated from magnitude estimations of single stimuli (km in Equation 3) should be the product of the input (k) and output (m) exponents separately estimated from magnitude estimations of "differences" using Equation 2. Curtis, Rule, and their associates conclude that they have found

[2] Rule and Curtis do not consider the additive constants in Equations 2 and 3 to be of deep theoretical significance and have not estimated these values from comparisons of two stimuli of equal value (Rule & Curtis, 1973b; Rule et al., 1974). In the experiments of Table 1, symmetric designs (including comparisons of a stimulus with itself) are used, and in this research, the observers almost always judge a zero physical difference to be "zero" for the "difference" task or "one" in the "ratio" task. Birnbaum and Veit (1974a, Footnote 1) describe graphical and statistical tests that are consistent with $b_R = 0$.

reasonable agreement between these two estimates.

In this theory, m represents the inverse of the psychophysical function for number. Attneave (1962) estimated m to be about 2.5 (1/.4) on the basis of evidence available at the time. Marks (1974) has noted that magnitude estimations can be approximated as a power function of category ratings using an exponent of 2.0, though his interpretation of this finding differs.

Ten experiments in which Equation 2 was fit to magnitude estimations of "differences" (Curtis, 1970; Curtis & Rule, 1972; Rule & Curtis, 1973a, 1973b; Rule, Curtis, & Markley, 1970; Rule, Laye, & Curtis, 1974) yielded estimates between 1.1 and 2.1, with an average value of $m = 1.47$ (see Rule & Curtis, in press, for a summary).

Scaling of number (Rule & Curtis, 1973a, in press) yielded an exponent for n of .63 (implying that $m = 1/n = 1.59$), which agrees well with estimates obtained by Rule (1972), Banks and Hill (1974), scales for number estimated from range–frequency theory (Birnbaum, 1974), scales estimated from the subtractive model (Rose & Birnbaum, 1975), and others. Rule and Curtis (in press) list estimates of n between .63 and .75, implying that m is between 1.33 and 1.59.

In sum, the theory of Curtis and Rule assumes that magnitude estimations are a power function of subjective value, with an exponent near 1.47. In agreement with Attneave's (1962) theory, the output function for magnitude estimation is considered to be the inverse of the psychophysical function for number.

A Two-Operation Theory

One can test the theory that subjects use two operations without assuming that the psychophysical functions for all stimulus continua are power functions. Indeed, the model can be tested using social stimuli, for which physical values are not defined. Equation 1 can be rewritten as follows:

$$R_{ij} = a_R(s_j/s_i)^m + b_R, \qquad (4)$$

where R_{ij} is the magnitude estimate of "ratio," and s_j and s_i are the subjective scale values (which may or may not be power functions of physical values). Assuming $b_R = 0$, it follows from Equation 4 that if one takes the log of both sides,

$$\log(R_{ij}) = m \log s_j - m \log s_i + \log a_R. \qquad (5)$$

Therefore, log magnitude estimations of "ratios" should fit a subtractive model, and when the log response is plotted as a function of the dividend stimulus, s_j, with a separate curve for each divisor, s_i, then the curves should be parallel. Such plots are usually roughly parallel (see, e.g., Birnbaum & Veit, 1974a, Figure 6), consistent with Equation 5.

The marginal mean log "ratio" judgment (averaging over levels of i) should be given by the following expression:

$$\frac{1}{r}\sum_{i=1}^{r} \log(R_{ij}) = m \log s_j + c, \qquad (6)$$

where r is the number of rows of the matrix, and c is a constant (independent of j).

Equation 6 implies that if "ratio" judgments are a power function of subjective ratios, then the marginal mean log "ratio" should be a logarithmic function of the scale value. If it is assumed that ratings of "differences" are linearly related to subjective differences (as evidenced by the near-parallelism in tests of the subtractive model), then the marginal mean "difference" ratings are a linear function of the scale values. In principle, even if it is only assumed that "differences" are a monotonic function of subjective intervals, it is possible to derive estimates of scale values for the subtractive model from nonmetric analysis. Therefore, assuming two operations and the principle of scale convergence (that scale values are independent of task), marginal mean log "ratios" should be a logarithmic function of subtractive model scale values derived from judgments of "differences."

Ratio–Difference Theory of Birnbaum and Veit

Birnbaum and Veit (1974a) proposed that "ratio" judgments may be computed by subtraction:

$$R_{ij} = J_R(s_j - s_i), \qquad (7)$$

where J_R is assumed to be strictly monotonic (it is usually found to be approximately exponential). Ratings of differences, D_{ij}, are also assumed to obey the subtractive model:

$$D_{ij} = J_D(s_j - s_i), \tag{8}$$

where D_{ij} represents the rating of the difference between stimuli j and i, and J_D is the strictly monotonic judgment function for ratings, which is typically estimated to be nearly linear.

If J_R is represented as an exponential and J_D as a linear function, Equations 7 and 8 can be specified, respectively,

$$R_{ij} = a \exp[b(s_j - s_i)]; \tag{9}$$

$$D_{ij} = c(s_j - s_i) + d. \tag{10}$$

Taking logarithms of both sides of Equation 9 implies

$$\ln R_{ij} = b(s_j - s_i) + a', \tag{11}$$

where $a' = \ln a$. Equation 11 shows that scale values can be estimated from the marginal means of the logarithms of the "ratio" responses. Equation 10 shows that estimates of the same scale values can be estimated from marginal mean "difference" ratings. The two sets of estimates are expected to be *linearly* related, according to this theory, in contrast with the *logarithmic* prediction of the two-operation theory.

Another implication of Equations 9 and 10 is that logarithms of "ratios" are predicted to be linearly related to "difference" ratings:

$$\ln R_{ij} = b' D_{ij} + d', \tag{12}$$

where $b' = b/c$ and $d' = a' - bd/c$. Equation 12 implies that "ratios" are exponentially related to "differences."

In the next section of this article, the two theories are compared by three tests applied to data for nine experiments. First, marginal means of log "ratios" are plotted against marginal mean "differences." The one-operation theory predicts that the relationship should be linear (Equations 11 and 12), whereas the two-operation theory predicts that the relationship should be logarithmic (Equation 6). Second, "ratios" are plotted against "differences." The one-operation theory predicts that the relationship should be monotonic, and the special case of Equation 12 predicts that it should be exponential. Third, the models are compared according to their relative success in reproducing the data.

A Reanalysis of Nine Experiments

Test 1: Marginal Means

Figure 1 shows the predicted relationship between marginal mean log "ratios" and marginal mean "differences" for the two theories. According to the theory of Birnbaum and Veit, assuming J_R in Equation 7 to be exponential, the relationship should be linear (Equation 11). According to the theory of two operations, however, the relationship should be logarithmic (Equation 6), and the domain of the log function will depend on m. Figure 1 shows that for values of m from 1 to 2.5, it should be possible to discriminate the two theories. However, as m approaches infinity, the theories become harder to distinguish. The average estimate, reported by Rule and Curtis (in press), is $m = 1.47$, which is intermediate among the values shown in Figure 1. Thus, if m is assumed to be a constant, the theories should be distinguishable on the basis of the test in Figure 1.[3]

Results: Marginal Means

Table 1 lists experiments for which the above test has been carried out. The table

[3] The predicted curve for the two-operation model was fit to the endpoints in Figures 1 and 2 as follows. First, the marginal mean log "ratios" were computed, the smallest marginal mean log was subtracted from all of them, and the antilog transformation was applied. This procedure theoretically sets the smallest estimated scale value (s_1) to 1.0, and the largest to $(s_k/s_1)^m$. Then the largest "ratio" scale value was raised to the .68 power (to correct for the presumed output function, $m = 1.47$), yielding s_k. Second, the "difference" task marginal means were linearly transformed to the range of values determined above from the marginal mean log "ratios" (i.e., from 1 to s_k). Finally, the log of each recalibrated "difference" scale value was multiplied by 1.47 to find the predicted curve shown in Figure 2 for the two-operation theory. In Figure 1 the same procedure was used, except both abscissa and ordinate have been calibrated to the same endpoints to highlight the effect of m. The predicted curves for the one-operation theory are straight lines through the endpoints.

shows that the studies have used a variety of stimulus continua and experimental procedures, in which the stimulus range, spacing, and manner of stimulus presentation have varied.[4]

Figure 2 shows the results of the first test for nine experiments. Straight lines have been drawn between the endpoints to represent the prediction of one-operation theory (Equation 11). Curves representing the theory of two operations (Equation 4) have also been fit through the endpoints; the value of $m = 1.47$ was used to generate the predictions (see Footnote 3). The points always fall below the logarithmic prediction of two-operation theory (Equation 6) and closer to the straight line prediction of one-operation theory (Equation 11).

The analyses of Figure 2 have also been carried out for the data of individuals, with the result that the vast majority of judges yielded data consistent with the group averages. For example, Birnbaum and Elmasian (1977) found that data for all eight judges had marginal mean log "ratios" that were close to a linear function of marginal mean "differences" (see open squares in Figure 2 of Birnbaum & Elmasian, 1977).

In sum, marginal means for nine experiments appear more consistent with one operation than with the theory of two operations, assuming the value of m to be between 1 and 2.5.

Test 2: Direct Ordinal and Metric Comparison

Because it is possible that two distinct rank orders can be transformed by separate weak monotonic functions to a degenerate solution (Rule & Curtis, 1980; Veit, 1980), it is preferable to compare the two rank orders directly, rather than ask if the two orders can be transformed to the same order. The left side of Figure 3 shows a comparison of actual ratios and differences for a 7×7, evenly spaced design. The scale values used for this example were the successive integers from 1 to 7. A ratio of 7:1 for the scale values was chosen for the example because the largest "ratio" judgment is often about 7. Actual ratios are plotted on the ordinate against actual differences, with a separate curve for each row of the factorial design (subtrahend or divisor). Thus, each curve represents s_j/c plotted against $s_j - c$, where c is the value of the subtrahend or divisor. The slope should be inversely related to the value of c, and the curves should cross where $s_j - c = 0$ (abscissa) and $s_j/c = 1$ (ordinate). For example, a difference of 4 could receive a ratio as large as 5 (5/1), or as small as 2.33 (7/3). Although 6/3 has the same ratio as 2/1, $6 - 3 > 2 - 1$. For constant differences, the ratio should approach 1 as the value of c increases. For constant ratios, the absolute difference should increase as c increases. The left of Figure 3 shows how different "ratios" and "differences" should be if the judge actually uses two operations and if magnitude estimations of "ratios" can be taken at face value ($m = 1.0$).

However, the theory of Attneave, Rule, and Curtis does not take magnitude estimates at face value. If the observer reports that one stimulus is "seven times" as intense as another, the subjective ratio is assumed to be only $7^{1/m}$. Assuming $m = 1.47$, the subjective ratio would be only 3.76. The predictions of this ratio theory were therefore calculated for a 7×7 design with subjective scale values spaced evenly between 1 and 3.76. Ratios were calculated according to Equation 4, and accordingly, each ratio was then raised to the 1.47 power. These values are plotted on the right of Figure 3 in the same way as on the left—the ordinate shows $(s_j/c)^{1.47}$ vs $2.17(s_j - c)$. The predictions, though less extreme, are quite distinct and follow the same general pattern as the ratios and differences on the left. Thus, even assuming that "ratios" are a power function of subjective ratios (using the exponent of Rule and Curtis), "ratios" and "differences" should not be monotonically related if judges use two operations, but should follow the pattern of predictions on the right of Figure 3.

On the other hand, if judges use only one

[4] Not listed in Table 1 are two experiments by Rose and Birnbaum (1975) that used the same response procedure for both tasks (a linemark) and therefore require a different analysis. Those experiments yielded conclusions consistent with the others.

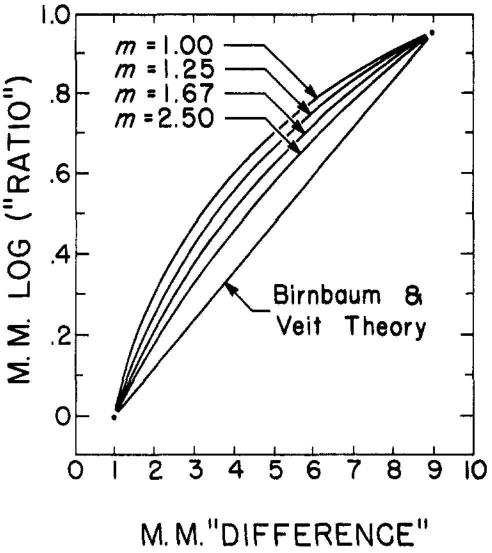

Figure 1. Theoretical relationships between marginal mean log "ratio" and marginal mean "difference." (One operation theory of Birnbaum and Veit [1974a] predicts a straight line; the theory that subjects use two operations predicts logarithmic curves. If the output function is a power function with exponent [m] greater than 1, then the domain of the log function is reduced. Rule and Curtis [in press] give an average value of $m = 1.47$; therefore, if this value of m is assumed to apply to "ratio" judgments, it should be possible to discriminate the two-operation theory from the one-operation theory.)

operation, according to Equations 7 and 8, then "ratios" and "differences" should be monotonically related, because $J_R^{-1}(R_{ij}) = s_j - s_i = J_D^{-1}(D_{ij})$. Therefore, $R_{ij} = J_R[J_D^{-1}(D_{ij})]$. Furthermore, if J_R is exponential and J_D is linear, then the relationship between "ratios" and "differences" will be exponential.

Results: Ordinal and Metric

Figure 4 plots the data as in Figure 3 for the nine experiments listed in Table 1. Figure 4 shows that "ratios" are very nearly a monotonic function of "differences," as predicted by the theory of Birnbaum and Veit (1974a). The data appear more consistent with the hypothesis that the points fall on the same monotonic function than with the predictions in Figure 3.

The relationship between "ratios" and "differences" is predicted to be exactly exponential when "ratios" are an exact exponential function of subjective differences and "differences" are linearly related to subjective differences. (In general, the judgment functions for both tasks are assumed to depend upon context effects such as the value of the modulus, the examples used to illustrate the scale, and the stimulus

Table 1
Experiments With "Ratio" and "Difference" Tasks

Reference	Dimension	Design	Stimulus range	Largest response example
Birnbaum & Veit (1974a)	Heaviness of lifted weight[a]	7 × 7	50–200 g (linear spacing)	400
Birnbaum & Elmasian (1977)	Loudness of 1000 Hz tones[b]	5 × 9	42–90 dB (log spacing)	400
Birnbaum (1978)	Darkness of dot patterns[c]	7 × 7	8–90 dots (log spacing)	800
Veit (1978, Experiment 1)	Darkness of gray papers[a,d]	7 × 7	.063–.572	800
Birnbaum & Mellers (1978)	Easterliness[a] of U.S. cities	7 × 7	San Francisco–Philadelphia	800
	Westerliness[a] of U.S. cities	7 × 7	Philadelphia–San Francisco	800
Hagerty & Birnbaum (1978)	Likableness of adjectives[a]	4 × 7	cruel–sincere	800
Elmasian & Birnbaum (Note 1)	Pitch of 78 dB (SPL) tones[b]	5 × 9	191–844 Hz (log spacing)	400
		5 × 9	191–3730 Hz (log spacing)	400

[a] Both stimuli in each pair were simultaneously presented; pairs were presented in random order.
[b] Stimuli were presented sequentially in restricted random orders.
[c] All stimuli were simultaneously presented.
[d] Between-subjects design.

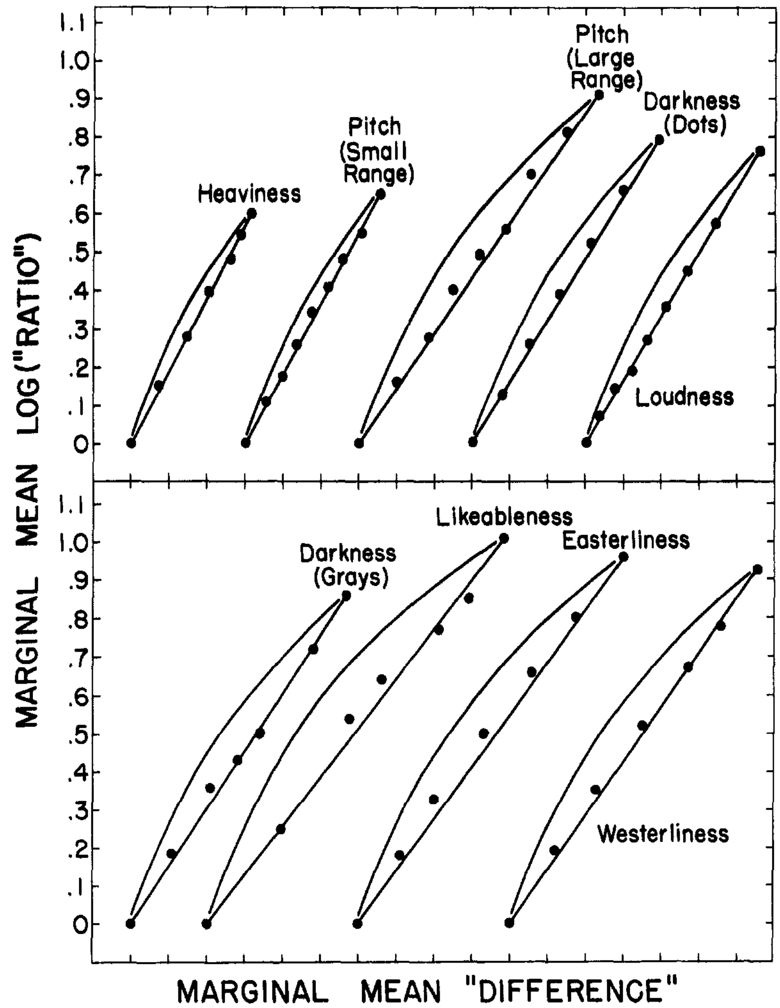

Figure 2. Marginal mean log "ratio" as a function of marginal mean "difference" for nine studies. (Straight lines drawn through the endpoints represent predictions of the one-operation theory of Birnbaum & Veit [1974a]. Curves drawn through endpoints show predictions of the two-operation theory assuming $m = 1.47$, which is the average value reported by Rule and Curtis. Abscissa scale has been linearly transformed for each experiment for comparability. A constant has been added to the marginal mean log "ratios" to make the smallest value zero in each case.)

distribution. Therefore, this relationship is expected to vary as a function of these variables). Exponential functions are drawn in Figure 4 through the highest point and the point (0,1) to permit assessment of the exponential relationship. Although there appear to be systematic deviations from the exponential curves (especially for "ratios" near one and "differences" near zero), these curves do capture the approximate form.[5]

Test 3: A Direct Comparison of Fit

Another way to test between theories is to compare their relative success in re-

[5] Curtis, Rule, and associates typically collect data for "differences" representing the upper triangle of a factorial matrix, excluding comparisons of a stimulus with itself (the diagonal). Unfortunately, their magnitude estimation experiments have not independently manipulated standard and comparison. Therefore, the key test shown in Figure 3 cannot be performed using their data.

producing the numerical data. Accordingly, to assess the theory of Birnbaum and Veit, the data for nine experiments were fit to the following models:

$$\hat{D}_{ij} = a_D(s_j - s_i) + b_D \quad (13)$$

and

$$\hat{R}_{ij} = a_R \exp(s_j - s_i) + b_R, \quad (14)$$

where $\hat{D}_{ij}$ and $\hat{R}_{ij}$ are the predicted "differences" and "ratios"; a_D, a_R, b_D, b_R, s_j, and s_i are parameters to be estimated from the data. The value of s_1 was fixed to 1.0, leaving 6 scale values to be estimated for each 7 × 7 design.

To represent the theory of two operations, Equation 14 was replaced with the ratio model,

$$\hat{R}_{ij} = a_R(s_j/s_i)^m + b_R. \quad (15)$$

Note that Equation 15 uses one additional parameter, m, in addition to the parameters required by Equation 14. To compare theories with an equal number of parameters, m was initially fixed to 1.47, the average estimate of Rule and Curtis (in press).

For each of the symmetric experiments described in Table 1, it was assumed that the scale values were independent of position. For those experiments (heaviness, darkness of grays, darkness of dots, easterliness, and westerliness), 10 parameters were estimated from each set of 98 (7 × 7 + 7 × 7) data values. For the asymmetric designs (loudness, pitch, and likeableness), different scale values were estimated for rows and for columns. For the likeableness experiment, different adjectives were used for row and column. In the loudness and pitch experiments, the tones were presented sequentially (the rows represent the first stimulus and the columns represent the second).

The index of fit was the sum of the proportions of total variance in the residuals added up over both "difference" and log "ratio" matrices. This index, L, was defined as follows:

$$L = \frac{\Sigma\Sigma(D_{ij} - \hat{D}_{ij})^2}{\Sigma\Sigma(D_{ij} - \bar{D})^2} + \frac{\Sigma\Sigma(r_{ij} - \hat{r}_{ij})^2}{\Sigma\Sigma(r_{ij} - \bar{r})^2}, \quad (16)$$

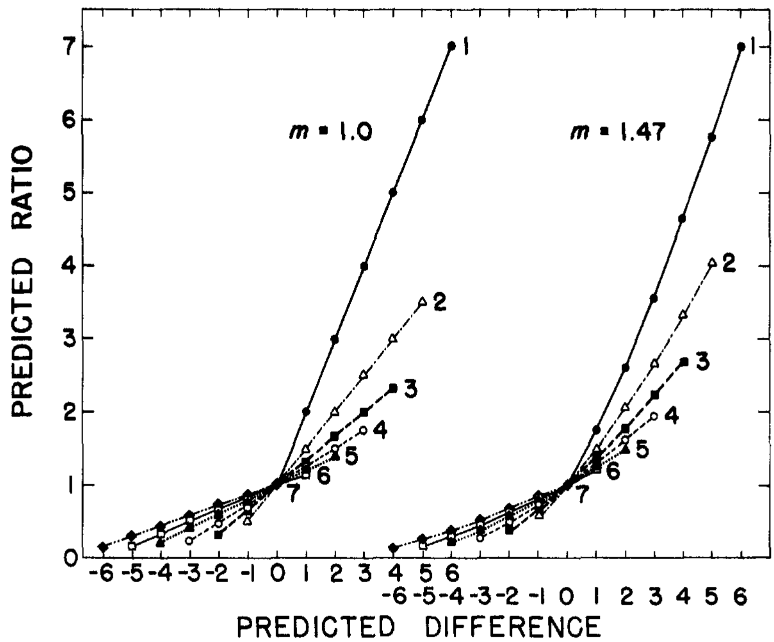

Figure 3. Left: Theoretical relationship between actual ratios and differences for an evenly spaced, 7 × 7 design using successive integers from 1 through 7. Right: Theoretical relationship for same design assuming "ratios" are a power function of subjective ratios with an exponent [m] of 1.47. (Note that ratios and differences are not functionally related: For a constant difference, the ratio approaches 1 as the divisor and subtrahend [curve parameter] increases.)

where L is the index to be minimized, D_{ij}, $\hat{D}_{ij}$, and $\bar{D}_{ij}$ are the ratings of "difference" between stimuli j and i, the predicted "difference," and the mean "difference," respectively. Because the standard errors for the "ratio" task vary directly with the mean "ratio," the deviations in the logs of the "ratios" ($r_{ij} = \log R_{ij}$, $\hat{r}_{ij} = \log \hat{R}_{ij}$, $\bar{r} =$

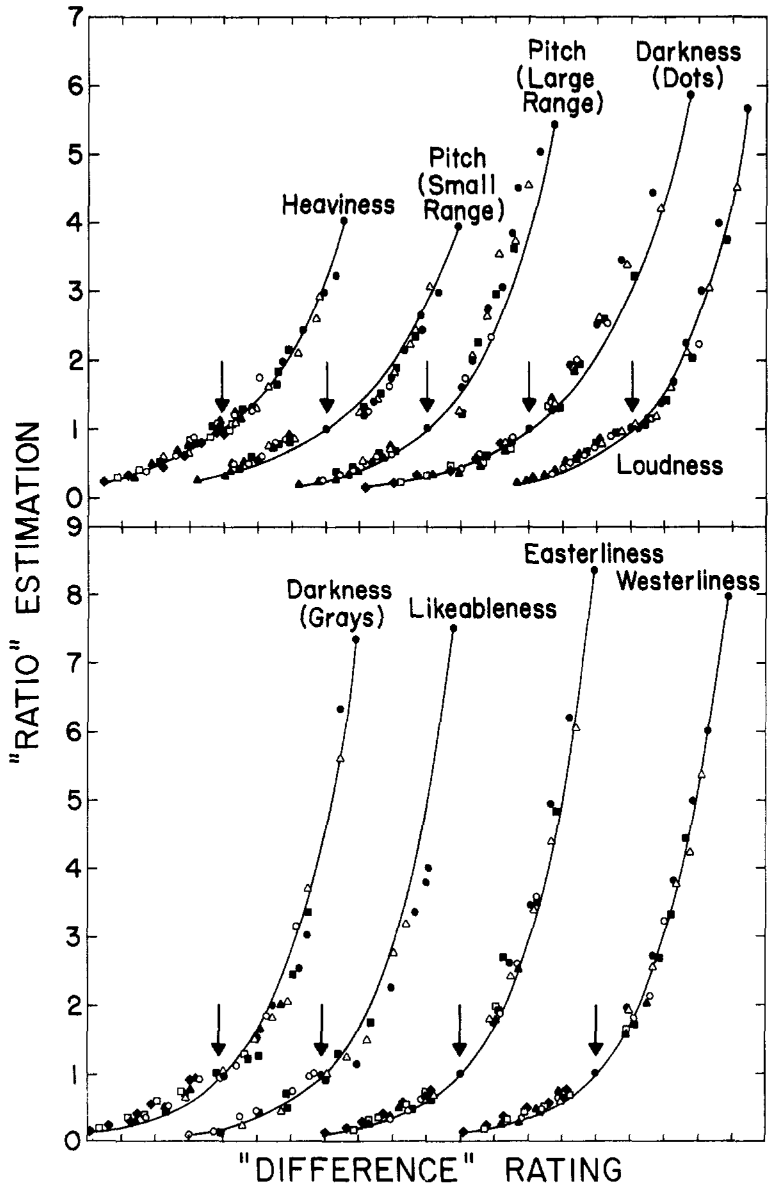

Figure 4. Magnitude estimations of "ratios" as a function of ratings of "differences" with a separate type of symbol for each "divisor," plotted as in Figure 3. (Curves show exponential functions drawn through the point [0,1] and the largest value for each experiment. Arrows show "zero difference" judgment; abscissa scales have been linearly transformed for each experiment for comparability. Results approximate a monotonic function more closely than they do the theoretical predictions in Figure 3, consistent with the hypothesis that "ratios" and "differences" are both governed by the same operation.)

mean log R) were minimized. A computer program was written to accomplish the minimization, utilizing the STEPIT subroutine (Chandler, 1969).

Results: Comparison of Fit

Equations 13 and 14, which model the one-operation theory, gave reasonable approximations to the data for all of the experiments. For example, for the darkness of grays experiment (Veit, 1978, Experiment 1), the residuals accounted for less than 1% of the total systematic variance for both "differences" and log "ratios," yielding an overall index of $L = .019$ summed over both sets of data. Six of the nine experiments had values of $L \leq .022$.

The one-operation model (Equations 13 and 14) gave a better fit to the data for all nine experiments than the model of two operations (Equations 13 and 15), with m fixed to the Rule and Curtis value of 1.47. The fit of the two-operation theory improves as m increases, but even with m fixed to 3.0, the Birnbaum-Veit theory provided a better fit for all but one experiment (likeableness). When m was free to vary, the estimated values of m were greater than 3 in all but one case, and the fit for the model of one operation was better than or equal to the fit of the model of two operations for all but one of the nine experiments.

If judges actually used only one operation, one would expect the fit of the two-operation theory to approach the fit of the one-operation theory as m increased (see Figure 1). When $m = 2.5$, actual differences in an equally spaced 7×7 design (with a largest "ratio" of 7 as in Figure 3) would *not* be a weak monotonic function of ratios calculated from Equation 15. However, when m exceeds 3.0, actual differences *are* a weak monotonic function of actual ratios for this design. Thus, in order for the theory of two operations to provide a reasonable fit to the data, it is necessary for it to make predictions that are essentially equivalent to those of the one-operation theory.

In sum, comparisons of fit indicate that to retain the theory that observers use two operations according to Equations 13 and 15 requires the conclusion that the exponent (of the output function for magnitude estimation of "ratios") exceeds 3. However, values beyond 3 appear unacceptable for the theory of two operations if m is supposed to be the reciprocal of the exponent for numerals, which Rule and Curtis (in press) find to be between .63 and .75.

Discussion

Comparisons and Contrasts

The psychophysical theories of Rule and Curtis (1980, in press) and Birnbaum and Veit (1974a), Birnbaum (1978), and Veit (1978) have five points of agreement:

1. Both approaches view psychophysical judgment as a composition of processes. For example, both assume that magnitude estimations can be represented as the composition

$$\mathrm{ME} = J_{\mathrm{M}}[H(\phi)],$$

where ME is the magnitude estimation of stimulus ϕ, H represents the psychophysical (input) function, and J_{M} represents the judgment (output) function.

2. Both agree that J_{M} is typically positively accelerated for magnitude estimation.

3. Both approaches agree that "difference" judgments can be represented by subtraction, $\mathrm{D}_{ij} = J_{\mathrm{D}}[s_j - s_i]$, where the output function, J_{D}, is presumed to depend on the dependent variable (magnitude estimation or category rating).

4. Both agree that the J function for category ratings is more nearly linear than the J function for magnitude estimation but that the exact form of both functions depends upon procedural details of the experiment.

5. Birnbaum and Veit (1974a) adopted the premise that the psychophysical function, H, is independent of the task to judge "differences" or "ratios." In the research of Rule and Curtis, it is assumed that H is independent of the task to estimate "differences" or "magnitudes."

The theories differ on three major questions, which are discussed below.

One operation or two? Rule and Curtis (1980) questioned the conclusions of Veit (1978) and suggested that observers may indeed use both ratio and subtractive operations for the corresponding instructions.

Birnbaum and Veit concluded that observers use only one operation for both tasks.

The data for the nine experiments listed in Table 1 appear consistent with the theory that observers compare two stimuli by the same process whether instructed to judge "differences" or "ratios." The data do *not* appear consistent with the theory that judges use both operations and that "ratio" judgments are a power function of subjective ratios with an exponent between 1 and 2.5. Although some small effect of instructions could exist, the difference due to task instructions is much smaller than predicted by the theory of two operations (Compare Figure 1 with Figure 2 and Figure 3 with Figure 4).

Is the output function for magnitude estimation the inverse psychophysical function for number? Rule and Curtis follow Attneave's (1962) theory that J^{-1} is the psychophysical function for number. They assume that J_M should be the same for magnitude estimations of "differences" and magnitude estimations of single stimuli. Birnbaum and Veit (1974a), on the other hand, assume that J_M represents a judgmental transformation that depends lawfully on context effects such as the examples used to illustrate the scale, the stimulus distribution, and other such details of experimental procedure.

The estimates of the psychophysical function for number from several studies using a variety of scaling techniques have been fairly consistent (Rule and Curtis, in press). However, estimation of m assuming two operations for the experiments of Table 1 would yield exponents that deviate from previous values by more than a factor of two.

The output function for magnitude estimation appears to depend on the range of examples used in the instructions to illustrate the magnitude estimation scale. Veit 1978, Experiment 1; Note 2) used the same stimuli (darkness of grays) with 2 different sets of examples ranging either from 25 to 400 or from 12.5 to 800 and found that the largest mean "ratio" was only 531 when examples ranged from 25 to 400, compared with 735 when the examples ranged from 12.5 to 800. For the data in Figure 4, the largest "ratio" estimation appears to be larger for experiments using 800 as the largest example (lower panel). The largest "ratio" judgment is smaller for heaviness, loudness, and pitch—experiments that used 400 as the largest example. Robinson (1976) reached similar conclusions for magnitude estimations of single stimuli. Apparently, the output function for magnitude estimation depends on the range of examples used to illustrate the scale in the instructions. Variations among individuals can often be attributed to variations in J (Rule & Curtis, 1977). It seems reasonable to suppose that the context effects in magnitude estimation described by Poulton (1968) can also be attributed to the output function, though most of the research cited by Poulton provides no basis for determining the locus of contextual effects.

Both schools of thought agree that the output functions depend upon experimental manipulations such as the value of the modulus, stimulus range, response range, and so on, but their interpretations differ. Rule (Note 3) tends to view these effects as perturbations that could increase or decrease the estimated exponent in different studies; he believes different estimates contain a true value that should on the average represent the psychophysical function for number. Birnbaum and Veit (1974a), however, view the J functions as labile and highly sensitive to variations in procedure. In their view, J is potentially predictable from principles of judgment but could attain a wide variety of forms, depending on the experimental conditions. Thus, Birnbaum and Veit view the J function as a predictable, but variable, transformation that should not be assumed to represent the subjective values of numbers.

Power or exponential? Rule and Curtis assume that the output function for magnitude estimation is a power function, whereas Birnbaum and Veit assume that it can be approximately exponential under certain experimental conditions.

The exponential output function for magnitude estimation explains how "ratio" judgments can fit a ratio model, even though the comparison operation is subtraction. If $R_{ij} = \exp(s_j - s_i)$, then $R_{ij} = \exp(s_j)/\exp(s_i) = s_j^*/s_i^*$, where $s^* = \exp(s)$. An

Table 2
Response Scales for "Difference" and "Ratio" Tasks

"Difference" rating scale		"Ratio" estimation scale
−3 = "A is much less than B"	←→	12.5 = "A is one eighth B"
−2 = "A is less than B"	←→	25 = "A is one fourth B"
−1 = "A is slightly less than B"	←→	50 = "A is one half B"
0 = "A equals B"	←→	100 = "A equals B"
1 = "A is slightly greater than B"	←→	200 = "A is twice B"
2 = "A is greater than B"	←→	400 = "A is four times B"
3 = "A is much greater than B"	←→	800 = "A is eight times B"

exponential output function explains why magnitude estimations of "averages" show bilinear divergence (Weiss, 1972) and explains otherwise contradictory results for the size–weight illusion. For further discussion, see Birnbaum and Veit (1974b) and Birnbaum (1978, pp. 52–53).

The following reasoning may help clarify conditions that produce the exponential judgment function for magnitude estimation. Magnitude estimation can be considered a type of category judgment in which the category names (numbers) used by the subject can be geometrically spaced (Birnbaum, 1978). The theory of Birnbaum and Veit assumes that the judge would be equally willing to use either of the response scales shown in Table 2, using corresponding categories (indicated by arrows) for the same stimulus pair. The examples in Table 2, if used in the instructions, can build in an exponential relationship between magnitude estimates and ratings. If the judge uses only one operation, subtraction, but uses the numbers next to the examples for magnitude estimations of "ratios," then the J_M function will be exactly exponential. If individuals choose different ranges of the response scale, but have identical subjective values, then judgments of "ratios" for different individuals will be related by power functions. Experimental manipulation of the range of the examples would be expected to influence J_M, such that results with different response ranges will be related by power functions.

Suppose for the moment that the subjective locations of U.S. cities can be represented by locations on a two-dimensional "mental map" with arbitrary origin (Birnbaum & Mellers, 1978). If so, then ratios are meaningless. Suppose the observer computes a directed distance, d, when asked to find the "ratio" of the easterliness of Philadelphia to that of San Francisco. Since this distance is the greatest that the observer is asked to report during the experiment, the observer uses the largest "ratio" response given in the instructions, 800 ("eight times"). When asked to judge the "ratio" of the easterliness of San Francisco to that of Philadelphia, the directed distance is $-d$. However, an equal and opposite subjective distance receives a reciprocal response, which in this case would be the smallest example in the response scale, 12.5 or "one eighth." Thus, three points that are psychologically equidistant, $-d$, 0, and d, produce responses of 12.5, 100, and 800, which are separated by equal ratios. In this fashion, equal subjective differences can produce equal ratios. Indeed, Birnbaum and Mellers (1978) found that "ratios" of easterliness and westerliness gave good fits to the ratio model. In other words, if the judge compares stimuli by subtraction, the instructions for magnitude estimation can induce an exponential judgment function, which causes the ratio model to give a good apparent fit.

Systematic deviations of fit from the ratio model applied to "ratio" judgments may be attributable to departures from the exponential judgment function rather than to any specific problem with the "ratio" task itself. The theory of Birnbaum and Veit (1974a) predicts that if the examples are not geometrically spaced, the exponential function will be violated, in which case the ratio model would not be expected to fit the raw data but would require monotonic transformation. Furthermore, the theory of Birnbaum

(1978) predicts that when different standards (or sets of standards) are used for different groups of subjects, the estimated exponent relating magnitude estimations to physical values should change, since the J function depends upon the stimulus and response distributions.

Beyond Torgerson's Indeterminacy

Torgerson (1961) concluded that if judges perceived only a single relation between a pair of stimuli, then it would be impossible to discover the nature of the single relation. Whichever representation for stimulus comparison was chosen would be a "decision not a discovery." However, Birnbaum (1978, 1979), Hagerty and Birnbaum (1978), and Veit (1978) have concluded that judges *do* use both ratio and subtractive operations, at least when they are instructed to judge "ratios of differences." Therefore, it is possible to discover whether the comparison process is best represented as a difference or ratio.

It appears that when subjects are instructed to judge "ratios" and "differences" of stimulus "differences," the data are consistent with the following:

$$\mathrm{RD} = J_{\mathrm{RD}}\,(\psi_{\mathrm{AB}}/\psi_{\mathrm{CD}}) \qquad (17)$$

and

$$\mathrm{DD} = J_{\mathrm{DD}}\,(\psi_{\mathrm{AB}} - \psi_{\mathrm{CD}}), \qquad (18)$$

where RD and DD are judgments of "ratios of differences" and "differences of differences," J_{RD} and J_{DD} are strictly monotonic judgment functions, and ψ_{AB} and ψ_{CD} represent subjective "differences." In principle, Equations 17 and 18 define a ratio scale of "differences" (ψ_{AB}) and thus permit one to *discover* the comparison function $\psi_{\mathrm{AB}} = s_{\mathrm{A}} \odot s_{\mathrm{B}}$, where $\odot$ is the comparison operation (Birnbaum, 1978). The data permit the following:

$$\psi_{\mathrm{AB}} = s_{\mathrm{A}} - s_{\mathrm{B}}, \qquad (19)$$

where s_{A} and s_{B} are subjective scale values. Therefore, Equation 19 is not arbitrary but implied by the data.

One could ask why observers should use two operations when the stimuli are *differences*, but not when the stimuli are *magnitudes*. It seems reasonable to suppose that humans possess the "mental machinery" to perform both operations but that they use subtraction whenever ratios are not meaningful. If the stimuli are inherently no more than an interval scale, like locations on a cognitive map, then ratios are not meaningful. However, differences between points always have a well-defined zero point ($x - y = 0$ if $x = y$) even when the original scale is only an interval scale. For example, it *is* meaningful to ask, "What is the ratio of the *distance* from San Francisco to Denver relative to the *distance* from San Francisco to Philadelphia?" Thus, ratio of differences is a meaningful operation on an interval scale. If visual length intervals are represented as differences between points, then perhaps observers can use two operations for "differences" and "ratios" of length. Parker, Schneider, and Kanow (1975) concluded that subjects use two operations for length comparisons.

In order to retain the ratio model for "ratio" judgments, it appears necessary either to suppose a very complex function for "ratios of differences" (Birnbaum, 1978; Veit, 1978), to abandon the principle of scale convergence, or to suppose that subjects "reinterpret" one of the four-stimulus tasks ("differences of differences") so that it can be represented by a log (ratio of ratios) model (Birnbaum, 1979; Eisler, 1978). Furthermore, the subtractive theory permits "ratios" and "differences" of easterliness and westerliness to be represented by a single cognitive map (Birnbaum & Mellers, 1978), and it permits one to retain a negatively accelerated psychophysical function for number (Rose & Birnbaum, 1975), whereas the ratio model requires two mental maps and a positively accelerated number function. For these reasons, Birnbaum (1978, 1979) concluded that the subtractive theory gives the most parsimonious account of the data.

Conclusion

The tests described in this article appear capable of distinguishing the theory that observers use two operations for "ratios" and "differences" from the theory that observers use only one operation for both tasks. Data for nine experiments using a variety of stimulus continua and experimental procedures strongly favor the theory that judges use only one operation whether instructed to judge "ratios" or "differences."

Reference Notes

1. Elmasian, R., & Birnbaum, M. H. *A harmonious note on pitch*. Unpublished manuscript, 1979.
2. Veit, C. T. *Effect of instructions on the output function for magnitude estimation*, Unpublished experiment, 1974.
3. Rule, S. J. Personal communication, 1979.

References

Attneave, F. Perception and related areas. In S. Koch (Ed.), *Psychology: A study of a science* (Vol. 4). New York: McGraw-Hill, 1962.

Banks, W. P., & Hill, D. K. The apparent magnitude of number scaled by random production. *Journal of Experimental Psychology*, 1974, *102*, 353-376. (Monograph)

Birnbaum, M. H. Using contextual effects to derive psychophysical scales. *Perception & Psychophysics*, 1974, *15*, 89-96.

Birnbaum, M. H. Differences and ratios in psychological measurement. In N. J. Castellan & F. Restle (Eds.), *Cognitive theory* (Vol. 3). Hillsdale N.J.: Erlbaum, 1978.

Birnbaum, M. H. Reply to Eisler: On the subtractive theory of stimulus comparison. *Perception & Psychophysics*, 1979, *25*, 150-156.

Birnbaum, M. H., & Elmasian, R. Loudness "ratios" and "differences" involve the same psychophysical operation. *Perception & Psychophysics*, 1977, *22*, 383-391.

Birnbaum, M. H., & Mellers, B. A. Measurement and the mental map. *Perception & Psychophysics*, 1978, *23*, 403-408.

Birnbaum, M. H., & Veit, C. T. Scale convergence as a criterion for rescaling: Information integration with difference, ratio, and averaging tasks. *Perception & Psychophysics*, 1974, *15*, 7-15. (a)

Birnbaum, M. H., & Veit, C. T. Scale-free tests of an additive model for the size-weight illusion. *Perception & Psychophysics*, 1974, *16*, 276-282. (b)

Chandler, J. P. STEPIT: Finds local minima of a smooth function of several parameters. *Behavioral Science*, 1969, *14*, 81-82.

Curtis, D. W. Magnitude estimations and category judgments of brightness and brightness intervals: A two-stage interpretation. *Journal of Experimental Psychology*, 1970, *83*, 201-208.

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

Curtis, D. W., & Rule, S. J. Magnitude judgments of brightness and brightness difference as a function of background reflectance. *Journal of Experimental Psychology*, 1972, *95*, 215-222.

Einstein, A. *Relativity: The special and the general theory*. New York: Crown, 1961.

Eisler, H. On the ability to estimate differences: A note on Birnbaum's subtractive model. *Perception & Psychophysics*, 1978, *24*, 185-189.

Hagerty, M., & Birnbaum, M. H. Nonmetric tests of ratio vs. subtractive theories of stimulus comparison. *Perception & Psychophysics*, 1978, *24*, 121-129.

Krantz, D. H., Luce, R. D., Suppes, P., & Tversky, A. *Foundations of Measurement*. New York: Academic Press, 1971.

Marks, L. E. On scales of sensation: Prolegomena to any future psychophysics that will be able to come forth as a science. *Perception & Psychophysics*, 1974, *16*, 358-376.

Parker, S., Schneider, B., & Kanow, G. Ratio scale measurement of the perceived lengths of lines. *Journal of Experimental Psychology: Human Perception and Performance*, 1975, *104*, 195-204.

Poulton, E. C. The new psychophysics: Six models for magnitude estimation. *Psychological Bulletin*, 1968, *69*, 1-19.

Robinson, G. H. Biasing power law exponents by magnitude estimation instructions. *Perception & Psychophysics*, 1976, *19*, 80-84.

Rose, B. J., & Birnbaum, M. H. Judgments of differences and ratios of numerals. *Perception & Psychophysics*, 1975, *18*, 194-200.

Rule, S. J. Comparisons of intervals between subjective numbers. *Perception & Psychophysics*, 1972, *11*, 97-98.

Rule, S. J., & Curtis, D. W. Conjoint scaling of subjective number and weight. *Journal of Experimental Psychology*, 1973, *97*, 305-309. (a)

Rule, S. J., & Curtis, D. W. Reevaluation of two models for judgments of perceptual intervals. *Perception & Psychophysics*, 1973, *14*, 433-436. (b)

Rule, S. J., & Curtis, D. W. Subject differences in input and output transformations from magnitude estimations of differences. *Acta Psychologica*, 1977, *41*, 61-65.

Rule, S. J., & Curtis, D. W. Ordinal properties of subjective ratios and differences: Comment on Veit. *Journal of Experimental Psychology: General*, 1980, *109*, 296-300.

Rule, S. J., & Curtis, D. W. Levels of sensory and judgmental processing: Strategies for the evaluation of a model. In B. Wegener (Ed.), *Social attitudes and psychophysical measurement*. Hillsdale, N.J.: Erlbaum, in press.

Rule, S. J., Curtis, D. W., & Markley, R. P. Input and output transformations from magnitude estimation. *Journal of Experimental Psychology*, 1970, *86*, 343-349.

Rule, S. J., Laye, R. C., & Curtis, D. W. Magnitude judgments and difference judgments of lightness and darkness. A two-stage analysis. *Journal Experimental Psychology*, 1974, *103*, 1108-1114.

Schneider, B., Parker, S., Farrell, G., & Kanow, G. The perceptual basis of loudness ratio judgments. *Perception & Psychophysics*, 1976, *19*, 309-320.

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

Torgerson, W. S. Distances and ratios in psychological scaling. *Acta Psychologica*, 1961, *19*, 201–205.
Veit, C. T. Ratio and subtractive processes in psychophysical judgment. *Journal of Experimental Psychology: General*, 1978, *107*, 81–107.
Veit, C. T. Analyzing "ratio" and "difference" judgments: A reply to Rule and Curtis. *Journal of Experimental Psychology General*: 1980, *109*, 301–303.
Weiss, D. J. Averaging: An empirical validity criterion for magnitude estimation. *Perception & Psychophysics*, 1972, *12*, 385–388.

Received May 22, 1979 ■