Perception & Psychophysics
1977, Vol. 22 (4), 383-391

Loudness "ratios" and "differences" involve the same psychophysical operation

MICHAEL H. BIRNBAUM
University of Illinois, Champaign, Illinois 61820

and

ROBERT ELMASIAN
University of California, San Diego, California 92093

Subjects judged both "ratios" of loudness and "differences" in loudness between pairs of tones that varied in intensity. The pairs were constructed from factorial designs, permitting separation of stimulus and response scaling for each subject. Ratings of "differences" and estimations of "ratios" were monotonically related, inconsistent with the hypothesis that subjects perform both subtractive and ratio operations on a common scale. Instead, the data suggest that both tasks involve the same psychophysical comparison operation with different response transformations. If the operation can be represented by the subtractive model, then category ratings involve a nearly linear transformation and magnitude estimations involve a nearly exponential transformation.

Torgerson (1961) theorized that subjects perceive only a single quantitative comparison between a pair of stimuli whether instructed to report a psychological "distance" or "ratio."[1] Torgerson based this hypothesis in part on Garner's (1954) finding that some subjects tended to make the same settings whether instructed to set a tone to bisect a loudness "interval" or to establish equal "ratios." The theory was also based on the observation that magnitude estimations are often an approximate exponential function of category ratings (Torgerson, 1960). However, with unifactor stimulus designs such as used in the early research (e.g., Stevens & Galanter, 1957; Torgerson, 1960), ratios and differences are necessarily monotonically related; hence, these early findings could not provide decisive ordinal tests of Torgerson's theory.

Recent research with factorial stimulus designs has provided stronger evidence for the theory that subjects do not distinguish "differences" from "ratios" (Birnbaum, 1977). Birnbaum and Veit (1974a) found that judgments of heaviness "ratios" and "differences" were monotonically related, contrary to the theory that subjects compute both ratios and differences on a common scale. Rose and Birnbaum (1975) found similar results for numerals, as did Veit (1977) for greyness of papers of varied reflectance.

The present paper investigates "ratios" and "differences" of loudness, using techniques of Birnbaum and Veit (1974a), applied to data for single subjects.

We thank Michael Hagerty, Colleen Surber, Barbara Mellers, Steven Stegner, and Clairice Veit for helpful comments on the manuscript. This research was supported in part by a grant from the Research Board of the University of Illinois, facilitated by Grant MH 15828 to the Center for Human Information Processing, University of California, San Diego. Requests for reprints should be sent to: Michael H. Birnbaum, Department of Psychology, University of Illinois, Champaign, Illinois 61820.

The Models

The *ratio* model under consideration here can be written:

$$R_{ij} = J_R(s_j/s_i), \tag{1}$$

where R_{ij} is the overt numerical "ratio" estimation for physical stimulus Φ_j to Φ_i, having sensations (scale values) s_j and s_i, and J_R is the monotonic judgmental transformation, relating overt magnitude estimations of "ratios" to subjective impressions of ratios.

The subtractive model can be written:

$$D_{ij} = J_D(s_j - s_i), \tag{2}$$

where D_{ij} is the rated "difference" between the stimuli, and J_D is the judgmental transformation relating overt ratings to subjective differences.

Metric Implications

If it is assumed that the judgment functions are linear, Equations 1 and 2 make distinct metric predictions for the data (Anderson, 1970; Birnbaum & Veit, 1974a). The ratio model, with the assumption that J_R is of the form: $R_{ij} = a(s_j/s_i)^b + c$, predicts a bilinear fan of curves. If the data are bilinear, the marginal means provide estimates of the scale values. If J_D is of the form: $D_{ij} = a(s_j - s_i) + c$, the subtractive model predicts that the responses from a factorial design should plot as a set of parallel curves. If the curves are parallel, scale values can be estimated from the marginal means.

Ordinal Indeterminacy

Although it would be convenient to assume that numerical judgments can be taken at face value, there are reasons to suspect that judgment functions may be nonlinear. For example, judgments change nonlinearly as a function of the responding procedure, the stimulus distribution, and other details of the experiment (Poulton, 1968; Stevens & Galanter, 1957). Birnbaum, Parducci, and Gifford (1971) have presented evidence that the stimulus distribution can affect the form of J (see also Birnbaum, 1974b). Therefore, it would be unwise to *assume* that numerical judgments are necessarily linearly related to subjective impressions. Once it is allowed that overt responses may require monotone rescaling, unifactor "ratio" scaling experiments only define the stimuli to an ordinal scale.

When rescaling is permitted, factorial designs allow ordinal tests of the models under investigation. Once the data from a suitable factorial experiment are found to be compatible with the model, the ordinal constraints are sufficient to solve for the scale values and the judgment function. However, even factorial designs leave an indeterminancy between the combination, or comparison, function and the judgment function.

When it is assumed only that the numerical judgments are a monotone function of psychological impressions (i.e., that J could be nonlinear), it would not be possible to discriminate ratio and subtractive models for a single set of data, since the ordinal requirements of both theories are the same (Krantz, Luce, Suppes, & Tversky, 1971; Krantz & Tversky, 1971). Data that are ordinally consistent with a ratio model could be monotonically transformed to fit a subtractive model, and vice versa. This indeterminacy cannot be resolved without additional constraints (Birnbaum, 1974a, 1977; Birnbaum & Veit, 1974a; Birnbaum, Note 1, Note 2).

Scale Convergence Criterion for Rescaling

Stimulus and response scale convergence criteria (Birnbaum, 1974a; Birnbaum & Veit, 1974a), together with data and theories of at least two situations involving the same stimuli, can be used to resolve some of the uncertainty in deciding whether or not to transform data to fit the hypothesized model. If the measurements of the stimuli are to have any meaning, they should allow one to interrelate different experimental relationships (see, e.g., Cliff, 1973). Thus, scale convergence (stimulus scale invariance) is an additional constraint that permits transformation while retaining the possibility of testing (i.e., rejecting) a set of models (Birnbaum, 1974a; Birnbaum & Veit, 1974a).

The scale convergence criterion permits transformations that fit hypothesized models *and* lead to invariant stimulus values. The study of two tasks with the scale convergence criterion permits ordinal tests among alternative explanations, which make different predictions.

Predictions

Assuming the models are ordinally consistent with the data, there are two simple possibilities: (a) two operations (both ratio and subtractive) with one scale, and (b) one operation with one scale.[2]

The theory that subjects compute both ratios and differences on a common scale implies that both Equations 1 and 2 must be consistent with the same scale. This theory implies that the two sets of data *will not* be monotonically related, but that the two rank orders of responses to pairs of stimuli will be interlocked (Krantz et al., 1971). For example, equal differences (e.g., 2-1, 3-2, 4-3) should correspond to ratios that approach one (e.g., 2/1, 3/2, 4/3) as the difference is moved up the scale. Similarly, equal ratios (2/1, 4/2, 8/4) correspond to more extreme differences (2-1, 4-2, 8-4) as the two values increase. If a single scale reproduces the orders in both factorial matrices, the scale values will have ratio scale uniqueness, since the ratio model defines scales unique to a power transformation and the subtractive model defines an interval scale (Krantz et al., 1971).

If subjects do not discriminate between "ratios" and "differences," however, then the *same* operation (e.g., either division or subtraction), applies to both kinds of judgments:

$$R_{ij} = J_R(s_j * s_i) \tag{3}$$

$$D_{ij} = J_D(s_j * s_i), \tag{4}$$

where * represents the comparison operation that is the same for both tasks, and J_R and J_D are the strictly monotonic judgment functions. Equation 4 implies $J_D^{-1}(D_{ij}) = s_j * s_i$; substituting in Equation 3, $R_{ij} = J_R[J_D^{-1}(D_{ij})]$. Since $J_R J_D^{-1}$ is a monotonic function, it follows that R_{ij} ("ratios") and D_{ij} ("differences") will be monotonically related. In sum, Torgerson's theory implies that there will be but one ordering of the pairs, whereas the theory of two operations implies two distinct orderings.

METHOD

Subjects performed two tasks: (a) They rated the "ratios" of loudness between tones, and (b) they estimated the "differences" in loudness.

For the "ratio" task, subjects estimated the "ratio of the subjective loudness of the second tone to the loudness of the first." The modulus (the value representing a ratio of unity) was designated "100," and printed examples were provided specifying that if the second tone seemed *one fourth as loud as the first*, the subject was to respond "25"; if it seemed *half as loud*, "50"; *twice as loud*, "200"; and *four times as loud*, "400." Instructions encouraged the subjects to feel free to use whatever numerical values best represented the "psychological ratios."

The "difference" task instructions required ratings of the "difference in loudness between the second tone and the first." Loudness "intervals" were rated on a 9-point scale, with category labels varying from 1 = *second tone is very, very much softer* to 9 = *second tone is very, very much louder*; 5 was designated as *equal*.

Stimuli, Design, and Apparatus

The stimuli were 1,000-Hz tone bursts that varied in sound pressure. The stimulus pairs were generated from a 5 by 9, First Tone by Second Tone, factorial design, in which the five levels of sound pressure for the first tone varied from 42 to 90 dB SPL in 12-dB steps (re: .0002 dyne/cm²); the nine levels of the second tone varied from 42 to 90 dB in 6-dB steps. The duration of each tone burst was 1 sec with an interstimulus interval of 1 sec.

The tone bursts were generated by a Wavetek Model 155 signal generator, controlled by a PDP-9 computer. The output of the Wavetek was bandpassed by a Krohn-Hite Model 355OR filter and presented monaurally to the subject through a TDH-39 earphone. Input to the earphone was measured with a Systron Donner Model 7014 frequency counter and a Bruel and Kjaer Type 2416 electronic voltmeter.

The subject, seated in a double-walled, Industrial Acoustics sound-treated chamber, responded by entering a number on an Elec-Trol KB10006 keyboard. Pressing the return key caused the computer to record the response and initiated the next trial after a .5-sec delay. The subject could repeat the pair of tones by pressing the space bar.

Subjects and Sessions

Eight paid members of the academic community of the University of California at San Diego served in four sessions on 4 different days. Each session was devoted to one task—either ratio (R) or difference (D). Two subjects performed the sessions in each of the following orders: RDRD, DRDR, RDDR, and DRRD. There were no discernible effects of day or task order.

Procedure

Each session began with instructions for both tasks, with emphasis on the task to be performed. The instructions emphasized the distinction between ratios and differences. They were told, for example, that although the differences between 2-1, 3-2, 4-3 are all the same, the ratios, 2/1, 3/2, 4/3, are all different. Although ratios of 2/1, 4/2, 8/4 are all equal, the differences increase. Ratios were described graphically as a partitioning of one line segment by another. Differences were illustrated graphically by cancellation of one line segment from another. It was noted that a change in order reverses the sign of a difference but produces a reciprocal ratio. The subjects were also required to complete 15 numerical calculations to demonstrate that they understood, at least intellectually, the distinction between ratios and differences. All subjects were able to perform these arithmetic tasks without error.

Each session consisted of a warm-up, followed by 10 replications of the 45 (5 × 9) stimulus pairs. Trial orders were separately randomized by the computer program for each subject and session, subject to the constraint that each block of 45 trials contained the entire 5 by 9 design.

RESULTS: METRIC ANALYSES

"Ratio" Task

Figure 1 plots mean "ratio" estimations (divided by the modulus, 100), with a separate panel for each subject. The highest curve in each panel represents "ratio" judgments of the second stimulus relative to the least intense first stimulus (42 dB). The lowest curve shows the results for the first stimulus of greatest intensity (90 dB). The spacing on the abscissa for the second stimulus, increasing from left to right, is not linear with the decibel scale. Rather, the abscissa spacing represents the marginal mean response for the second stimulus, averaged over levels of the first. Assuming a ratio model, spacing the abscissa according to the marginal means should cause all of the curves to plot as straight lines.

The panels for individual subjects are arranged in decreasing order of their largest mean "ratio" judgment. Note that each division of the ordinate and abscissa represents 1 unit, but the range of the upper panels is 10 whereas the lower panels is 4 (note that the graphical size of the unit is larger in the lower panels), so that subject F.L. has a smaller largest mean "ratio" estimation than J.S.

According to the ratio model, Equation 1, with an appropriate judgment function (J) for the magnitude estimation response, the curves in each panel should form a bilinear fan of diverging lines. Every subject's data show a divergent interaction that is approximately bilinear. However, there do appear to be two small departures from bilinearity. First, when the

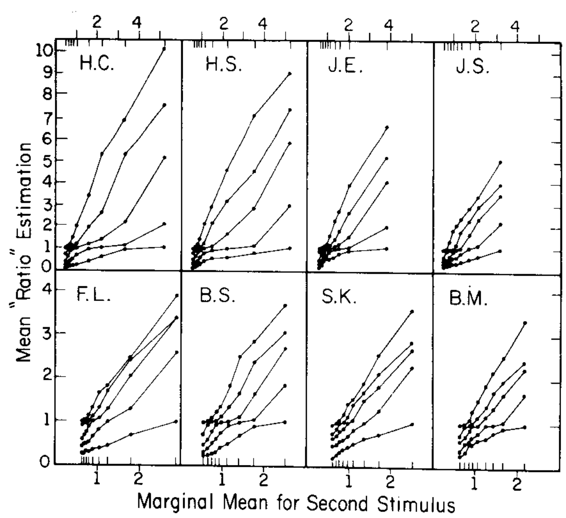

Figure 1. Mean "ratio" estimations as a function of marginal mean for the second stimulus (numerator) with a separate curve for each level of the first stimulus (denominator). Levels of the first stimulus are 42 (highest curve), 54, 66, 78, and 90 dB (lowest curve). Each panel represents the results for a different subject. Note that the ordinate and abscissa ranges are smaller in the lower panels than in the upper.

tones differ by small amounts (6 dB), the response is closer to a "ratio" of one than would be predicted by the ratio model; that is, the curves tend to have lower slopes as they cross the ordinate value of 1. Second, the curves for intermediate values of the first stimulus are curved upwards relative to the others. These two effects are similar to previous findings for magnitude estimation, and may represent monotonic response "biases" in the J function.

The subjects show large individual differences in their response ranges. Subject H.C. reports that the sensation "ratio" of 90 to 42 dB exceeds 10:1; J.S. estimates it as near 5:1; and subject B.M. estimates this "ratio" to be a little above 3:1. Although the subjects differ greatly in the values of their numerical judgments, they do appear similar with respect to the *pattern* of judgments. One interpretation of this individual variation in the response range is that it represents differences in the judgmental transformation, J_R, rather than "true" differences in the individuals' psychophysical functions.

"Difference" Task

The solid points in Figure 2 represents mean ratings of loudness "differences" as a function of the marginal means for the second stimulus, $\overline{D}_{.j}$. Each solid curve represents a different value of the first stimulus. Since the subjects rated second minus first, the slopes are positive. The upper curve is for the 42-dB tone; the lowest curve represents the 90-dB first tone. Each panel plots results for a different subject, as in Figure 1.

The subtractive model, Equation 2, with the assumption of a linear judgment function, predicts that the solid curves should be parallel. The curves do appear roughly parallel. However, there are deviations. Small (6-dB) differences are judged "equal" (a response of "5") too often. The slopes become too flat in the region of 5 on the ordinate, an effect that is analogous to that in Figure 1. Moderate differences appear to be exaggerated; note that the curves become steeper for 12-dB differences which tend to be judged "6" or "4"—i.e., "slightly louder" or "slightly softer." As shown below, these small deviations may represent judgmental factors (nonlinearity in the J function) that can be removed by monotone transformation.

Ratio-Difference Scale Convergence Fails

The open circles in the upper part of each panel of Figure 2 plot the marginal means for the "ratio" estimations ($\overline{R}_{.j}$) as a function of the marginal means for the "difference" ratings ($\overline{D}_{.j}$). Assuming both models were valid (with linear J functions in Equations 1 and 2), the marginal means are estimates of the scale values; therefore, the open circles should fall on a straight line (Birnbaum & Veit, 1974a). A

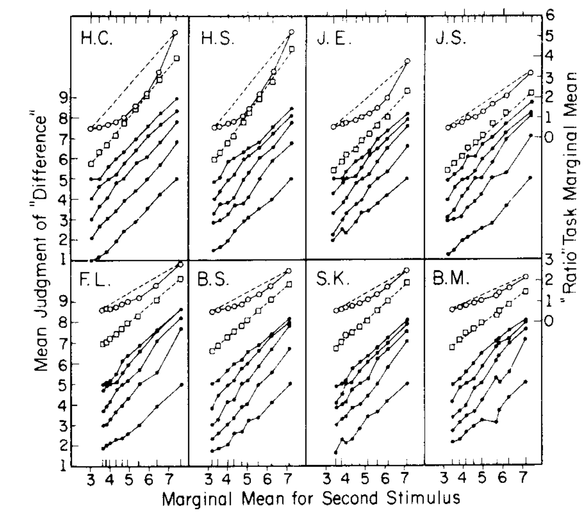

Figure 2. Solid points represent mean "difference" judgments, plotted as a function of marginal means for second stimulus with a separate curve for each level of the first stimulus. Open circles are marginal mean "ratio" estimations for the second stimulus, plotted against marginal mean "differences." Open squares are marginal mean logs of "ratio" estimations for the second stimulus.

dashed line has been drawn between the two end points to aid visual inspection of the nonlinearity. Instead of being linear, the curves appear more nearly exponential. This violation of scale convergence calls these assumptions (two operations on a common scale with linear judgment functions) into question.

Scale Convergence Consistent with One Operation

Suppose there are two operations, but magnitude estimations induce a nonlinear, power function bias in the judgment function for the "ratio" task. It would follow that the ratio model would still fit, but the marginal means (scale values) would differ from the subtractive model scales by a linear function of a power function. This interpretation would allow a nonlinear relationship between the marginal means. [This hypothesis is equivalent to a suggestion by Marks (1974) that scale values for "ratios" are the square of the scale values for "differences," since $R = (s_j/s_i)^b$ implies $R = s_j^b/s_i^b$.]

Logarithmic transformation of the "ratio" judgments yields $\log(R_{ij}) = \log(s_j/s_i)^b = b(\log s_j - \log s_i)$. Hence, this theory implies that log "ratio" judgments should be parallel, with marginal means logarithmically related to the marginal means (scale values) for the "difference" task.

On the other hand, suppose subjects make "ratio" estimations by computing subjective differences, then reporting a response exponentially related to their subjective impressions. "Ratios" would show a bilinear pattern (as in Figure 1), since $\exp(s_j - s_i)$

$= \exp(s_j)/\exp(s_i)$. This theory predicts that logarithmic transformation should undo the effects of the exponential J transformation. Hence, marginal mean logs should be a linear function, not a logarithmic function, of "difference" task marginal means.

The marginal mean logs are plotted in Figure 2 (open squares) as a function of marginal mean "difference" ratings. Dashed lines have again been drawn between the endpoints to permit assessment of linearity. The open squares appear very nearly linear, consistent with the results of Birnbaum and Veit (1974a). Assuming there is one scale, these results favor the interpretation that there is but one operation.

RESULTS: ORDINAL ANALYSES

The data were monotonically rescaled to fit the subtractive model by means of MONANOVA (Kruskal & Carmone, 1969), a computer program for nonmetric scaling. A separate rescaling was performed for each subject, task, and session. Each set of 450 judgments was rescaled to fit the model:

$$J_T^{-1}(T_{ijk}) = s_j - s_i + e_k, \qquad (5)$$

where J_T^{-1} is the strictly monotonic inverse of the judgment function for task T; T_{ijk} is the response ("ratio" or "difference") to the kth repetition of the presentation of stimulus pair ij; s_i and s_j are the scale values of the first and second stimuli, respectively; and e_k is an additive effect of repetition block.

This method of transformation is a conservative procedure, since the transformation that reduces deviations from the model is constrained to simultaneously reduce the error term for the model. Analyses of variance of the transformed scores were performed for each subject, task, and session (32 in all). Following transformation, the First Stimulus by Second Stimulus interactions (5 by 9) were reduced to less than 8/10ths of 1% of the total sum of squares in all cases and less than 3/10ths of 1% in 20 of the 32 cases. The Trial Blocks by First Stimulus by Second Stimulus interaction (10 by 5 by 9), which might be interpreted as an error term, was simultaneously reduced to less than 1% of the total sum of squares in 19 of 32 cases; less than 3.1% in all cases.[4]

Figure 3 shows the mean transformed response for each subject averaged over repetitions and sessions. Open circles represent mean transformed "ratio" estimations. Solid points connected by lines represent mean transformed "difference" judgments as a function of sound pressure level of the second tone with a separate curve for each level of the first tone.

The success of the transformations should be apparent in the parallelism of the open circles (rescaled "ratios") and solid points and curves (rescaled "differences") in Figures 3 and 4. This

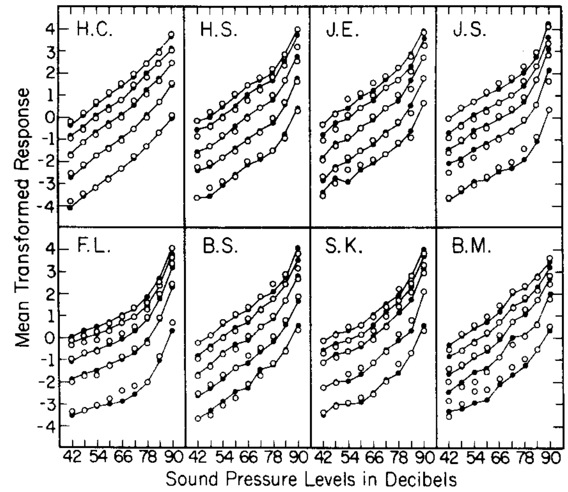

Figure 3. Mean transformed response as a function of the sound pressure level of the second tone with a separate curve for each level of the first. Solid points connected by lines are for "difference" ratings; open circles are for "ratio" estimations. To the extent that the orderings of points are the same (i.e., that open circles fall on the curves), scale convergence would suggest that *one* operation underlies both "ratio" and "difference" judgments.

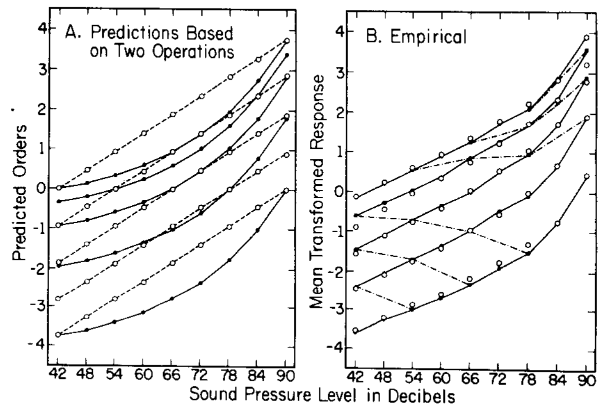

Figure 4. (A) Predictions based on the theory that subjects compute both ratios and differences on a common scale, plotted as in Figure 3. Open circles and dashed lines represent transformed predicted ratios; solid points and lines represent predicted differences. Hypothetical predictions assume scale values are a power function of physical energy; therefore, transformed ratios are linear. The distinction between ratios and differences does not depend on the psychophysical function, since no transformation of the abscissa could cause the circles and points to coincide. (B) Mean transformed response, as in Figure 3, averaged across subjects. Solid points connected by lines represent rescaled "difference" judgments; circles represent rescaled "ratio" judgments. Parallelism implies ordinal consistency with either a ratio or subtractive model. Broken lines connect different pairs with the same physical ratio (decibel difference). Power law ratio model (or logarithmic law for subtractive model) implies that broken lines should be horizontal and solid lines should be linear. Equivalence of orders suggests that *one* operation underlies both tasks.

parallelism indicates that the data for either task (analyzed separately) are ordinally consistent with either Equation 1 or Equation 2.

Only One Rank Order

The theory that the subjects use the same comparison operation for "differences" and "ratios" but different judgment functions implies that the rank orders of the "ratio" and "difference" judgments will be the same.

Figure 3 shows that for each subject, rescaled "ratios" and "differences" are nearly identical; i.e., that the open circles fall close to the curves. This result indicates that the rank orders for the two tasks are not consistent with the hypothesis of two operations on one scale (Equations 1 *and* 2). Instead, they are consistent with the hypothesis of one operation on one scale (Equations 3 and 4).

Figure 4 summarizes hypothetical predictions based on two operations (Panel A) and the empirical results (Panel B). In Figure 4A, the open circles connected by dashed lines represent rescaled computed ratios. The solid points connected by solid lines represent hypothetical rescaled "difference" judgments. Although the particular predicted values were calculated using sensory scales that are a power function of physical sound pressure, it should be clear that the general pattern of predicted differences between ratio and difference tasks does not depend on the power function assumption.

Predictions were made by either subtracting these scale values or taking their ratio, transforming to parallelism, and linearly adjusting them to the same scale. The assumption that psychophysical loudness is a power function would imply linearity for the ratio estimations. The ratio predictions are independent of the value of the exponent, since the rank order of ratios is invariant under power transformations.

Both sets of transformed predicted scores in Figure 4 are parallel. An illusion makes the curves appear to converge to the right. The reader should convince himself of the parallelism of the curves in Figure 4A by physical length measurement. Although this visual illusion may have some interest in its own right, it makes the transformed curves appear to converge, and must be taken into consideration in examining the visual appearance of parallelism in Figures 3 and 4.

The empirical means of the transformed scores, averaged over subjects, are shown in Figure 4B. Three aspects of the data are noteworthy. First, the averaged values are representative of the single subject transformed values (Figure 3).

Second, the mean transformed data are very nearly parallel for both tasks. The average absolute discrepancy from parallelism in Figure 4B is only .056 for the "ratio" data and only .032 for the "difference" task data. Hence, the ordinal properties of each set of data appear compatible with either a subtractive or a ratio model.

Third, and most important, the transformed values for the "ratio" and "difference" tasks appear nearly identical. The transformed data do not resemble the predictions (Figure 4A) based on two operations. Assuming one psychological scale of loudness, the similarity of the orders of the points suggests that the same operation underlies both tasks. Marginal means of the transformed scores (estimates of scale value) were nearly identical for both tasks. Thus, the criterion of scale convergence is consistent with the hypothesis of one operation.

DISCUSSION

The data for each task, analyzed separately, are approximately consistent with the numerical (metric) predictions of the model defined by the task given to the subject. However, scales derived from the two models are inconsistent. When the criterion of scale convergence is assumed, it is possible to reject the hypothesis that both operations, ratio and difference, are being computed on common scale values. It is also possible to reject the hypothesis that there are two scales related by a linear function of a power function.

The results of these experiments are consistent with Torgerson's (1961) hypothesis and with the interpretations of Birnbaum and Veit (1974a): the same comparison operation appears to underlie both tasks. To account for the differences in the numerical data, it suffices to assume different monotone judgment transformations for magnitude estimations of "ratios" and category ratings of "differences." Similar conclusions have also been reached by Beck and Shaw (1967), Birnbaum (1977), Rose and Birnbaum (1975), Schneider, Parker, Kanow, and Farrell (1976), and Veit (1977).

Schneider et al. (1976) performed nonmetric scaling analysis on loudness "ratio" judgments and fit the resulting subtractive model scale values to a power function of stimulus intensity. The fitted exponent was close in value to one reported in a previous experiment of loudness "differences" (Parker & Schneider, 1974). Schneider et al. concluded on the basis of the close match that loudness "ratios" and "differences" are governed by the same perceptual process. Although their conclusions were based on a questionable comparison of power function exponents (across experiments that differed in other respects besides instructions), it is reassuring to see independent lines of investigation reaching similar conclusions.

Psychophysical "Laws"

There is a tradition in psychophysics to fit curves with power functions. This can always be done, of course. However, just because one can approximate some data by means of a particular function does not mean that the function gains any theoretical status as an explanation. In a satirical article, Sue Doe (Nihm, 1976) argued that sensation is a polynomial function of physical value, since polynomial functions fit psychophysical data better than power functions. Nihm's Sue Doe law seems unlikely to gain acceptance, but it serves to illustrate the problem with considering the "psychophysical law" as a problem in curve-fitting.

The substantive laws by which stimuli are combined or compared deserve greater attention than the function relating scale values to physical measures. Sensation scales are not obtained by "operational definition," but through their role in fitting models. It is only with respect to these theories that scales of sensation have any meaning.

Violations of the Power "Law"

The scales defined in the present research depend on whether the subtractive or ratio interpretation is assumed. It is interesting that even under the assumption of a ratio model, the data are incompatible with the power function. Schneider et al. (1976) reached the same conclusion, using it as an argument against the ratio model interpretation. This seems ironic, since the power function was originally based on "ratio" scaling.

The power function implies that the curves in Figures 3 and 4B be linear. This follows from the ratio model, since if $R_{ij} = J_R(s_j/s_i)$ and $s = a\Phi^b$, then $J_R^{-1}(R) = a\Phi_j^b/a\Phi_i^b = (\Phi_j/\Phi_i)^b$; hence $\log [J_R^{-1}(R)] = b[\log \Phi_j - \log \Phi_i]$. Therefore, equal physical ratios should be judged equal, and rescaled log-"ratios" should be a linear function of log physical values. Instead, Figures 3 and 4B show that the curves for both tasks are nonlinearly related to the log physical values and that equal physical ratios (broken lines in Figure 4B) become more extreme as a given decibel difference is moved up the scale.

Of course, power functions could still be forced on the data, and exponents extimated.[5] These inappropriate exponents are intentionally *not* reported here for fear that they may end up in the table of a review article. It also seems inappropriate to force power function fits on the subtractive model scale values at present, since the qualitative constraints tested for the ratio model are not available for the subtractive model. In addition, fitting a power function to the subtractive model scales allows the estimation of an extra parameter, which could favor the possibly spurious conclusion that the power function was an appropriate representation of the subtractive model scales.

Epitaph for a Prolegomenon?

Marks (1974) proposed a prerequisite to any "future psychophysics that will be able to come forth as a science." The promulgated prolegomenon is a pair of power functions for "ratio" and "difference" tasks. Marks (1974) noted that if power functions are fit to data from these tasks, the power function exponents are greater for "ratio" tasks than for "interval" tasks. (See also Stevens, 1971.) Sensory scales are presumed to be power functions of physical magnitude, but with different exponents for different tasks.

If two scales were power functions of intensity, however, then one scale must be a power function of the other. These assumptions predict different orders for the two tasks. In fact, the predictions of Figure 4A would be unaffected by any power transformation of the ratio scales and any linear transformation of the difference scales. The finding that the two tasks produce the same order rules out the theory that there are two different power functions. To argue that ratio and subtractive operations underlie the same order would require the interpretation that the two scales are exponentially related. Rather than assume that there are *two* scales which just happen to be exponentially related and *two* operations, it would seem simpler to explain ordinal identity by postulating that there is one scale of sensation and one comparison operation.

Marks' (1977) stage theory of psychophysics concedes that both "ratios" and "differences" are computed by subtraction. The new theory allows different transformations of psychological value in different stages to account for differences among scales derived from additive models of loudness summation, subtractive models of "difference" judgments, and scales obtained by taking magnitude estimations at face value. Marks' new theory remains consistent with a variety of data, but seems complicated, since it allows so many different processes to intervene between stimulus and response.

Concluding Comments

Although the present data suggest that the same operation underlies judgments of both "ratios" and "differences," they do not provide any basis for preferring the ratio or subtractive model as a representation of this single operation. If the subtractive model is assumed to represent both tasks, the parallelism of the curves in Figure 3 imply that J_D is nearly linear, whereas the near-bilinearity in Figure 2 would imply that J_R is nearly exponential. However, if the ratio model is assumed, J_D would be logarithmic and J_R a linear function of a power function. Scale values derived assuming the ratio model would be an exponential function of those estimated from the subtractive model. Either theory could equally well reproduce the present data. Torgerson (1961) thought that the decision between these theories could not be made on the basis of empirical evidence.

An empirical basis can be made for preferring the subtractive representation (Birnbaum, 1977; Veit, 1977). Rose and Birnbaum (1975) asked judges to divide a line segment to represent "ratios" or "differences" between numerals. The responses were virtually identical for both tasks, consistent with the premise that only one operation was involved. The ratio model [and a related $s_i/(s_i+s_j)$ ratio model] implied a positively accelerated psychophysical function for numerals. The subtractive model, in contrast, yielded scale values that agree with values for numerals derived from Parducci's range-frequency theory (Birnbaum, 1974b) and with other scales of number obtained with a variety of techniques (Rule & Curtis, 1973).

Veit (1977) applied the scale-free approach (Birnbaum, 1974a; Birnbaum & Veit, 1974b) to the ratio-difference problem. She presented judges with four gray chips and asked them to judge the "ratio" of the "difference" in the darkness between two stimuli relative to the "difference" between two others. The data were consistent with a ratio of differences model. Since both a ratio and a subtractive operation are present in this model, the ordinal constraints preclude rescaling to another simple model. The scale values for darkness derived from this model were linearly related to scale values obtained from the subtractive representation of two-stimulus "ratios" or "differences," supporting the subtractive interpretation of these tasks.

Birnbaum (1977) noted that judgments obtained with "difference of differences," "difference of ratios," and "ratio of ratios" instructions are all consistent with a difference of differences model. "ratios of differences" and "differences of differences," however, are not monotonically related, but instead seem to be consistent with two operations on a common scale of differences. These two orders define a ratio scale of pair relationships, which, in turn, are consistent with the subtractive representation. The subtractive theory allows a coherent interpretation of all six tasks (including simple "ratios" and "differences") with a common scale.

Thus, "ratios" and "differences" appear to induce two different operations when the stimuli are themselves subjective pair intervals. Since an interval will always have a well-defined zero point (when two subjective values are equal, the interval is zero), it is tempting to hypothesize that two operations will be operative only for continua with well-defined subjective zero points (Birnbaum, 1977). Perhaps line lengths can be thought of as intervals, or distances, between locations. Hence, ratio and subtractive processes might both be available for length intervals (Parker, Schneider, & Kanow, 1975). However, for such continua as darkness of gray papers and dot patterns, magnitudes of numerals, heaviness of lifted weight, likeableness of adjectives, and now loudness of tones, it appears that only one operation is needed to represent both "ratio" and "difference" judgments.

REFERENCE NOTES

1. Birnbaum, M. H. *Impression formation: Difference judgments as a basis for response rescaling.* Paper presented at the meeting of the Western Psychological Association, San Francisco, April 1971.

2. Birnbaum, M. H. *When models fail and scales converge: Additional constraints for functional measurement.* Paper presented at the Conference on Stimulus Classification and Judgment, Los Angeles, May 1972.

REFERENCES

Anderson, N. H. Functional measurement and psychophysical judgment. *Psychological Review*, 1970, **77**, 153-170.

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

Birnbaum, M. H. The nonadditivity of personality impressions. *Journal of Experimental Psychology Monograph*, 1974, **102**, 543-561. (a)

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

Birnbaum, M. H. Differences and ratios in psychological measurement. In N. J. Castellan & F. Restle (Eds.), *Cognitive theory* (Vol. 3). Hillsdale, N.J: Erlbaum, 1977, in press.

Birnbaum, M. H., Parducci, A., & Gifford, R. K. Contextual effects in information integration. *Journal of Experimental Psychology*, 1971, **88**, 158-170.

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)

Cliff, N. Scaling. *Annual Review of Psychology*, 1973, **24**, 473-506.

Garner, W. R. A technique and a scale for loudness measurement. *Journal of the Acoustical Society of America*, 1954, **26**, 73-88.

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

Krantz, D. H., & Tversky, A. Conjoint measurement analysis of composition rules in psychology. *Psychological Review*, 1971, **78**, 151-169.

Kruskal, J. B., & Carmone, F. J. MONANOVA: A FORTRAN-IV program for monotone analysis of variance. *Behavioral Science*, 1969, **14**, 165-166.

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.

Marks, L. E. Phonion. In N. J. Castellan & F. Restle (Eds.), *Cognitive theory* (Vol. 3). Hillsdale, N.J: Erlbaum, 1977.

Nihm, S. D. Polynomial law of sensation. *American Psychologist*, 1976, **31**, 808-809.

Parker, S., & Schneider, B. Nonmetric scaling of loudness and pitch using similarity and difference estimates. *Perception & Psychophysics*, 1974, **15**, 238-242.

PARKER, S., SCHNEIDER, B., & KANOW, G. Ratio scale measurement of the perceived lengths of lines. *Journal of Experimental Psychology: Perception and Performance*, 1975, **104**, 195-204.

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

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

RULE, S. J., & CURTIS, D. W. Conjoint scaling of subjective number and weight. *Journal of Experimental Psychology*, 1973, **97**, 305-309.

SCHNEIDER, B., PARKER, S., KANOW, G., & FARRELL, 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.

STEVENS, S. S., & GALANTER, E. H. Ratio scales and category scales for a dozen perceptual continua. *Journal of Experimental Psychology*, 1957, **54**, 377-411.

TEGHTSOONIAN, R. On the exponents in Stevens' law and the constant in Ekman's law. *Psychological Review*, 1971, **78**, 71-80.

TORGERSON, W. S. Quantitative judgment scales. In H. Gulliksen & S. Messick (Eds.), *Psychological scaling: Theory and applications*. New York: Wiley, 1960.

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*, 1977, in press.

NOTES

1. Quotation marks are used throughout for instructions to judge "differences" or "ratios," or for judgments obtained with such instructions. Quotes are not used to denote theoretical statements about models or actual ratios and differences.

2. It is logically possible to obtain data for each task that are compatible with the models, yet the scale values from the two tasks would not have any simple relationship. The data for one or both tasks could also violate the ordinal requirements of one or both models. These possibilities have been discussed in greater detail by Birnbaum (1977), Birnbaum and Veit (1974a), and Krantz et al. (1971).

3. If the data can be rescaled to fit the subtractive model, then they are also ordinally compatible with the ratio model. Hence, the rescaling to parallelism tests either model.

4. Marginal means for the five levels of the first stimulus were a nearly linear function of the corresponding values of the second stimulus.

5. It is interesting to note that the log range of magnitude estimations of "ratios" is 1.43, yielding a (log response range)/(log stimulus range) ratio of .30. These values are typical of results obtained with magnitude estimations of loudness of single tones (Teghtsoonian, 1971). In the present case, however, this ratio *cannot* be interpreted as a power function exponent, and should raise serious questions about the interpretation of this index for single stimulus experiments.

(Received for publication December 24, 1976;
revision accepted July 11, 1977.)