Journal of Experimental Psychology:
Human Perception and Performance
1984, Vol. 10, No. 2, 216-230

Weight of Evidence Supports One Operation for "Ratios" and "Differences" of Heaviness

Barbara A. Mellers and Diane M. Davis
University of California, Berkeley

Michael H. Birnbaum
University of Illinois, Urbana-Champaign

This paper investigates an apparent contradiction between recent studies of "ratios" and "differences" of heaviness. Birnbaum and Veit (1974) found a single rank order for judgments in the two tasks, whereas Rule, Curtis, and Mullin (1981), who used a different stimulus set, procedure, and experimental design, reported two orders. To investigate the cause of this discrepancy, the present study manipulated the experimental design using the same stimuli and procedure as Rule et al. (1981). In one experiment (within-subject designs), each subject judged all combinations of the standard and comparison stimulus; in the other experiment (between-subjects designs) each subject received only one standard, and different groups of subjects were given different standards. "Ratios" and "differences" of heaviness were monotonically related for the majority of subjects who judged all combinations of standards and comparisons. Variations in the modulus and response examples did not affect the rank order of "ratios" within subjects. These results suggest that the contradiction in results is due to the difference in experimental design rather than differences in stimuli or procedure. In the between-subjects designs, the rank order of the "ratio" judgments depended on the standards and examples. Both previous and present results are consistent with the theory that subjects use one operation, subtraction, for both tasks and that the judgment function varies with between-subjects manipulations of the standard, examples, and modulus.

Early work in psychological scaling relied on the assumption that subjects do what they are instructed to do. For example, if a subject, instructed to judge the "ratio" of heaviness of two weights, called the "ratio" 2:1, it was assumed that the sensations of the two weights also stood in the same ratio. However, because scales based on instructions to judge "differences" and "ratios" are nonlinearly related (Garner, 1954; Stevens and Galanter, 1957), it is no longer assumed that subjects follow instructions. To distinguish between instructions and theories, quotation marks are used for subjects' judgments of "ratios" and "differences," but not for theoretical ratios and differences.

Although there were no empirical grounds to prefer one procedure over the other, Stevens (1957, 1971) argued that category ratings are biased. His arguments were based on a preference for magnitude estimation and the untested assertion that "ratio" judgments yield a ratio scale of sensation. A large literature in psychophysics grew from the dubious assumptions that "ratio" judgments can be represented by a ratio model and provide ratio scales. These assumptions led to Steven's power law. For commentary from different viewpoints, see Ekman and Sjoberg (1966), Krantz, Luce, Suppes, and Tversky (1971), and Shepard (1981).

Torgerson (1961) noted that the contradictions in the scales could be understood if it were assumed that subjects are insensitive to the mathematical distinctions imposed by difference or ratio instructions and use a single comparison operation. Unfortunately, Torgerson's one-operation hypothesis could not be well tested using unifactor designs popular 20 years ago, because x-c and x/c are always monotonically related when x is a variable and c is a constant. However, the one-operation

The authors thank Stanley Rule for comments.

Requests for reprints should be sent to Barbara Mellers, Department of Psychology, University of California, Berkeley, California 94720.

theory can be tested when x and c are independently manipulated because x/c and x-c are generally not monotonically related; for example, 6-4 > 2-1, but 6/4 < 2/1.

Two- Versus One-Operation Theories

The theory that subjects use both ratio and difference operations as instructed can be written as follows:

$$R_{ij} = J^*(s_j/t_i) \qquad (1)$$

$$D_{ij} = J(s_j - t_i), \qquad (2)$$

where R_{ij} is the "ratio" estimation; D_{ij} is the "difference" rating; s_j and t_i are the subjective values of the first and second stimulus, respectively; J^* and J are strictly monotonic judgment functions relating subjective impressions to responses for "ratios" and "differences," respectively.

Figure 1 illustrates the predictions of the two-operation theory of Equations 1 and 2. The stimuli are evenly spaced on the cube root of physical weight, and the scale values are based on a power function of weight, $s_i = \phi_i^{0.72}$, as in Rule et al. (1981). Differences between scale values, $(s_j - s_i)$, are plotted on the left; logs of ratios, (s_j/s_i), are plotted on the right. (Because the log function is a monotonic transformation, log ratios have the same rank order as ratios.) Predictions are shown as a function of the level of Weight A (first stimulus) with a separate curve for each level of Weight B (second stimulus). Dashed lines, connecting stimulus pairs separated by the same number of stimulus levels, highlight the predicted change in rank order between ratios and differences. Notice that the dashed lines diverge to the right for differences and converge to the right for ratios. In summary, Figure 1 shows that the rank order of "ratios" and "differences" should be quite distinct if subjects use two operations as in Equations 1 and 2.

The theory that subjects use one operation, subtraction, regardless of instructions to judge "ratios" or "differences" can be written

$$R_{ij} = J^*(s_j - t_i) \qquad (3)$$

$$D_{ij} = J(s_j - t_i), \qquad (4)$$

where the terms are defined as in Equations 1 and 2. According to the one-operation theory, the rank order of judged "ratios" and "differences" should be the same, because $R_{ij} = J^*(J^{-1}(D_{ij}))$. Unlike Figure 1, "ratio" and "difference" judgments should be monotonically related.

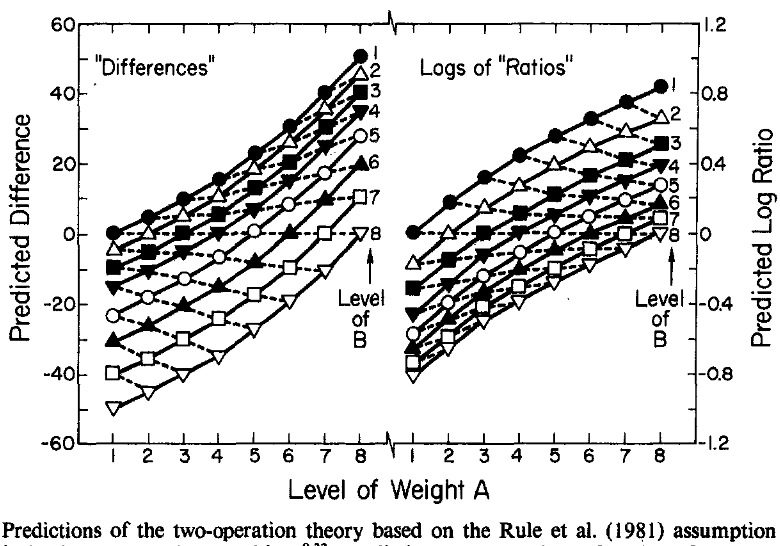

Figure 1. Predictions of the two-operation theory based on the Rule et al. (1981) assumption that $s_i = \phi_i^{.72}$. (Physical values are evenly spaced in $\phi_i^{0.33}$. Predictions are plotted as a function of the level of A with a separate curve for each level of B. Differences, A − B, are plotted on the left and logs of ratios, log(A/B), are plotted on the right. Different symbols are used for each level of the second stimulus (Weight B). Dashed lines connect points associated with stimulus pairs that are separated by the same number of stimulus levels to highlight the different rank orders in the two panels.)

Recent research has shown that when subjects are asked to judge "ratios" and "differences" of stimulus pairs constructed from factorial designs, judgments from the two tasks are monotonically related (Birnbaum, 1982). This result has been replicated with several psychological continua, including loudness, pitch, heaviness, darkness of grays, likeableness of adjectives, and others (Birnbaum, 1978, 1980, 1982; Birnbaum & Elmasian, 1977; Birnbaum & Mellers, 1978; Birnbaum & Veit, 1974; Hagerty & Birnbaum, 1978; Schneider, Parker, Farrel, & Kanow, 1976; Veit, 1978).

The finding that both tasks yield the same rank order can be explained by the theory that subjects use one-comparison operation for the two tasks. Evidence suggests that when subjects are asked to judge either "ratios" or "differences," the single operation is best described by a subtractive model, as in Equations 3 and 4. See Birnbaum (1982) for a discussion of this argument.

Apparent Contradiction in Results for Heaviness

Birnbaum and Veit (1974) obtained judgments of "ratios" and "differences" of heaviness and found data consistent with a one-operation theory. However, Rule et al. (1981) recently reported two distinct orderings for "ratios" and "differences" of heaviness, in contrast with previous results.

Rule et al. (1981) expressed at least three reservations about previous research on the ratio–difference question. First, they argued that the one-operation theory may be appropriate for "metathetic" continua such as position but not for "prothetic" continua such as heaviness or loudness. Second, they argued that it might be difficult to detect two operations unless the stimulus range and spacing were properly selected. Third, they argued that when each subject judges all stimulus combinations, "ratio" judgments may be biased in such a way that "ratios" and "differences" are monotonically related.

Rule et al. (1981) designed their experiments to increase the chances of detecting two rank orders, given their concerns noted above. They used different stimuli from those of Birnbaum and Veit (1974) and a different experimental design that employed different groups of subjects for each standard in the "ratio" task. Because the experiments of Rule et al. (1981) and Birnbaum and Veit (1974) differed on at least 13 variables, it is impossible to know how to interpret the discrepancy in the results.[1] The purpose of the present experiments is to test hypotheses concerning the apparent contradiction between the two studies.

A Hypothesis for the Discrepancy

Rule et al. (1981) implicitly assumed that different groups of subjects who receive different standards have the same J^* functions. However, Mellers and Birnbaum (1982) found that different groups of subjects who receive different standards, stimuli, and/or examples can yield "ratio" judgments that violate Equation 1. Mellers and Birnbaum (1982) proposed that variations of stimuli, standards, and/or examples in between-subjects designs produce different J^* functions for different groups of subjects. They suggested that magnitude estimations obtained from between-subjects designs can be represented by a subtractive model with different J^* functions, that is,

$$R_{ij} = J_i^*(s_j - t_i), \tag{5}$$

where J_i^* is a different monotonic function for each group of subjects receiving different conditions (e.g., standards), t_i. Although different from Equation 1, Equation 5 is consistent with the results of Rule et al. (1981).

Given the results of Mellers and Birnbaum (1982), it seems reasonable to hypothesize that the apparent contradiction between the results of Birnbaum and Veit (1974) and Rule et al. (1981) may be due to the use of different experimental designs in the "ratio" tasks. To test

[1] Rule et al. (1981) used stimuli that had a different minimum, a different range, and different spacing. The lifting procedure of Rule et al. (1981) involved lifting weights that were unseen (rather than seen) by means of the wrist (rather than the arm), by grasping a ring (rather than a cylinder). The weights were lifted sequentially (rather than simultaneously) by one hand (rather than two). The experimental design used a different group of subjects for each standard (rather than each subject's receiving all standards). A triangular (rather than a factorial) design was used for "differences." Instructions specified a modulus of 1 instead of 100, and examples were not geometrically spaced (vs. geometrically spaced). Subjects made "absolute" rather than algebraic "difference" judgments.

this hypothesis, the present study uses two experimental designs for "ratio" judgments (between-subjects vs. within-subject variations of standards). In Experiment 1, the stimuli and lifting procedure of Rule et al. (1981) are employed, but each subject receives all possible combinations of the standard and comparison stimuli. In Experiment 2, different groups of subjects receive different standards for "ratio" judgments, as in Rule et al. (1981). In addition, different response examples were used in both studies to test whether changes in the examples affect the J^* functions, as in Equation 5.

If the apparent contradiction in the results is due to the different range and spacing of the stimuli or the different modulus and lifting procedure, then Experiment 1, which uses the same stimuli and procedure as Rule et al. (1981), should replicate their findings: The data should be consistent with the two-operation theory, as in Figure 1. On the other hand, if the results are due to the use of different experimental designs (between-subjects vs. within-subject) for standards in the "ratio" judgments, then Experiment 1 should replicate the results of Birnbaum and Veit (1974): "Ratios" and "differences" should be monotonically related; furthermore, the between-subjects conditions of Experiment 2 should be consistent with the interpretation that changes occur in the J^* function, as in Equation 5.

Method

Overview

There were two experiments with different subjects in each. In Experiment 1, judges (subjects) estimated "differences" and "ratios" of the heaviness of pairs of weights. Each subject judged both "differences" and "ratios." There were three conditions in Experiment 1 (with different subjects in each) that used different values for the modulus and/or range of examples in the "ratio" task, but each subject judged all combinations of the standard and comparison stimulus. In Experiment 2, judges estimated only "ratios" of heaviness. Four conditions in Experiment 2 were constructed from a 2 × 2 between-subjects factorial design of standard by range of examples so that each subject received only one standard and one set of response examples but several comparison stimuli.

Apparatus and Stimuli

The apparatus for both experiments was similar to that used by Rule et al. (1981). The judge sat at a table and lifted weights by raising two metal rings 1 in. in diameter labeled A (on the left) and B (on the right). The rings were attached to nylon cords that passed through holes in the table. Weights were attached to the nylon cords under the table and were not visible to the judge. In both experiments, the eight weight values in grams were identical to those used by Rule et al. (1981): 20, 35.04, 56.18, 89.48, 120.97, 166.67, 222.66, and 290, including the weight of the line, hook, and rings.

Procedure

In both experiments, judges were tested individually. They were instructed to rest their arm on the table and use their wrist to lift each weight. The two weights were lifted alternately with the preferred hand. The sequence of lifts for a pair of weights began with the stimulus on the left (A), followed by the stimulus on the right (B). Judges were told they could lift the weights as many times as they liked, but always in the same order, left then right.

Experiment 1

Instructions. In the difference task, subjects rated the "difference" in heaviness between Weight A and B, (A − B), using the following scale: *8* = A is very, very much heavier than B, *6* = A is very much heavier than B, *4* = A is heavier than B, *2* = A is slightly heavier than B, *0* = A and B are equally heavy, *−2* = B is slightly heavier than A, *−4* = B is heavier than A, *−6* = B is very much heavier than A, and *−8* = B is very, very much heavier than A. Subjects could use any integers between +8 and −8.

In the "ratio" tasks, subjects were instructed to estimate the "ratio" of heaviness of Weight A to Weight B, (A/B). In different conditions, different examples were provided in the "ratio" instructions. In the "32" condition, the examples were: *32* = A is 32 times as heavy as B, *16* = A is 16 times as heavy as B, *8* = A is 8 times as heavy as B, *4* = A is 4 times as heavy as B, *2* = A is 2 times as heavy as B, *1* = A and B are equally heavy, *0.5* = A is ½ as heavy as B, *0.25* = A is ¼ as heavy as B, *0.125* = A is ⅛ as heavy as B, *0.0625* = A is $\frac{1}{16}$ as heavy as B, and *0.031* = A is $\frac{1}{32}$ as heavy as B. This range was selected in an attempt to replicate the response range obtained by Rule et al. (1981).

In the "4" condition, the examples were a subset of those in the "32" condition and ranged from *4* to *0.25*. These examples resemble those used in Rule et al. (1981), according to Rule (personal communication, 1981).

In the "800" condition, the modulus was 100 instead of 1.0. The examples read: *800* = A is 8 times as heavy as B; *400* = A is 4 times as heavy as B; *200* = A is 2 times as heavy as B; *100* = A and B are equally heavy; *50* = A is ½ times as heavy as B; *25* = A is ¼ as heavy as B; *12.5* = A is ⅛ as heavy as B. These examples are similar to those used previously by Birnbaum and his colleagues (see Birnbaum, 1980).

In all three conditions, subjects were encouraged to use numbers in between or more extreme than the examples provided.

Design. The same stimulus design was used for "ratio" and "difference" tasks in all three conditions of Experiment 1. The eight values of Weight A (on the left) were paired with the same eight values of Weight B (on the right) in an 8 × 8 factorial design.

Each judge served in six 1-hr sessions. During a session the judge completed 10 warm-up trials for one task followed by two randomly ordered replications of the 64 trials for that task. Subjects repeated each task six times over the six sessions. Task order was alternated throughout the sessions. Half of the judges received the tasks in the order RDRDRD; the other half received the opposite task order, DRDRDR. Each block of 64 trials was presented in a different order by shuffling a deck of 64 cards that represented the 64 (8 × 8) cells in the design. Cards were shuffled for every block of 64 trials given to each subject.

Experiment 2

Instructions. In Experiment 2, subjects were instructed to estimate the "ratio" of the heaviness of Weight A to Weight B, (A/B). There were two different sets of examples that were identical to the "32" and "4" conditions of Experiment 1, and different subjects received different examples.

Design. Each subject received eight values of Weight A, the comparison stimulus (on the left), but only one value of Weight B, the standard (on the right). Different subjects received different standards, 56.18 or 84.48 g. Subjects judged each stimulus five times. Thus, there were four groups of subjects who received one of the two response examples combined with one of the two standards.

Each judge served in one ½-hr session. During a session, the judge completed four warm-up trials followed by five replications of the eight trials for the "ratio" task. Each block of eight stimuli was presented in a different random order for every replication for each subject by shuffling a deck of eight cards.

Judges

Subjects were 56 members of the academic community of the University of California, Berkeley. The 24 judges in Experiment 1 (8 different subjects in each condition) were either acquaintances of the experimenters or recruited by an advertisement. They received $20.00 for participation in the six sessions. The 32 judges in Experiment 2 were either acquaintances of the experimenters or undergraduates who participated for credit in an introductory psychology course. There were eight judges in each of the four conditions. In both experiments, judges were naive with respect to the purpose of the study and had no previous experience with such experiments.

Results

Experiment 1: Within-Subjects Standards

One or two orderings? Graphs similar to Figure 1 were drawn separately for each subject in the three conditions of Experiment 1. The majority of subjects showed data inconsistent with the predictions of Figure 1. Instead, for most subjects, the rank order of "difference" judgments was quite similar to that of "ratios." Only 3 subjects out of 24 appeared to show different orderings in the two tasks, and their data will be discussed separately later.

Means for all but these 3 subjects are shown in Figure 2, plotted as in Figure 1. "Differences" are shown on the left; logs of "ratios" are shown on the right. ("Ratio" judgments from the "800" condition were divided by 100 prior to log averaging.) A different symbol is used for each level of the second stimulus (Weight B), as in Figure 1.

On the left, dashed lines connecting stimulus

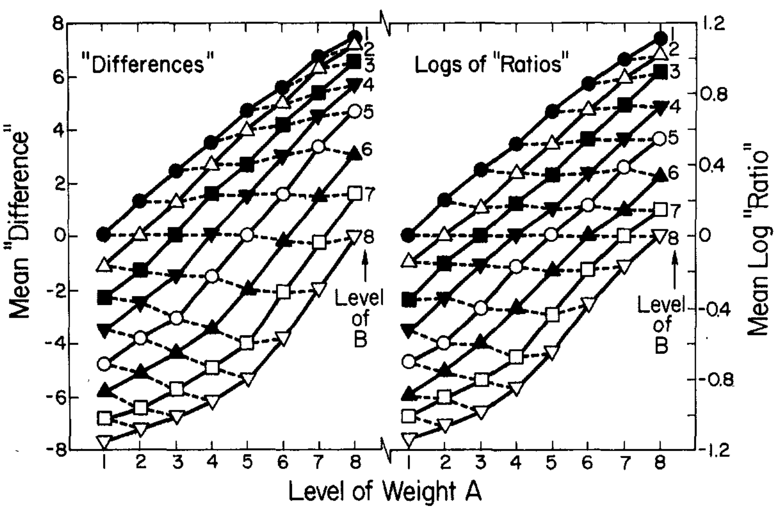

Figure 2. Mean "differences" (on the left) and mean logs of "ratios" (on the right), as plotted in Figure 1. (Notice that the dashed lines, connecting stimulus pairs separated by the same number of stimuli, diverge to the right for "differences." Dashed lines for "ratios" resemble "differences," unlike Figure 1.)

pairs separated by the same number of stimulus levels for "differences" diverge slightly to the right, that is, the "difference" between the pairs becomes more extreme as one moves up the scale. Similarly on the right, dashed lines connecting stimulus pairs for logs of "ratios" also diverge slightly to the right. Within-subject "ratios" differ from the between-subjects "ratios" of Rule et al. (1981); the data do not conform to the two-operation predictions of Figure 1.[2]

Furthermore, the data in Figure 2 permit the rejection of any power function for heaviness if the ratio model is assumed. A power function implies that judged "ratios" will have the same rank order as ratios of physical value. However, the rank order of judged "ratios" in all three conditions differs from the rank order of actual ratios. Compare the dashed lines in the right-hand panels of Figures 1 and 2. Birnbaum and Elmasian (1977) also found that the ratio model and power function are incompatible for loudness.

The data of 3 subjects appeared to show two rank orders. One of the 3 subjects had a large order effect in his "difference" data only, such that when a heavy weight (7 or 8, for example) was presented twice, the second presentation was judged heavier. When a light weight was presented, the second was judged lighter, as though the memory of the first weight regressed toward the mean. Another subject had a small trend in the direction of two different orders. The third subject showed a clear pattern of two orders consistent with Figure 1. This subject (Subject 18 in Table 1) voluntarily reported during the course of the experiment that he decided there were exactly eight weights. He informed the experimenter that he had assigned the numbers 1 through 8 to the successive weights, calculated a numerical difference or ratio depending on the task, and gave the computed number as his response. This strategy would yield data consistent with two operations. This subject's data were more consistent with the two-operation theory than were the data of the other two subjects (Subjects 5 and 20). Birnbaum and Mellers (1978) also found 2 subjects who employed the same strategy by writing down digits from 1 to 7 to represent the stimuli and calculating their responses.

Effects of the modulus and range of examples. The three conditions of Experiment 1 yielded similar rank orders for both tasks. There appeared to be no systematic effect of the modulus and response examples on the rank order of "ratios." However, the "ratio" examples did affect the response range. The last column of Table 1 shows the largest "ratio" response for each subject in the three conditions. The largest geometric mean "ratios" (290 g:20 g) were 11, 25, and 831 (actually 8.3, because the modulus was 100) in the "4," "32," and "800" conditions, respectively. The smallest mean responses were 0.08, 0.04, and 11.7 (.117) compared with the smallest examples of ".25," ".031," and "12.5."

Figure 3 plots predicted ratios versus differences for the two-operation theory from Figure 1, using the same symbols as in Figures 1 and 2. Predicted ratios were calculated to match the response range obtained by Rule et al. (1981). The figure shows $(s_j/s_i)^{1.71}$ versus $s_j - s_i$. Note that if subjects use two operations, the curve of solid points should fall strictly above the curve of open triangles, which should be above the curve of solid squares, and so forth. The figure shows best the order for ratios greater than 1. The predicted change in the order for ratios less than 1 is as great as for ratios above 1, but can be seen more clearly with logs of ratios as shown in the lower half of Figure 1.

Figure 4 shows geometric mean "ratios" plotted against arithmetic mean "differences" separately for each of the three conditions (excluding the 3 subjects mentioned earlier), using the same symbols as in Figures 1, 2, and 3. "Ratios" for the "4" and "32" conditions should be read against the left hand ordinate, whereas those for the "800" condition should be read against the scale on the right. "Difference" ratings are shifted along the abcissa to separate the conditions. Solid curves rep-

[2] To check whether any systematic differences occur when "ratios" and "differences" are obtained between subjects, responses for the first session of Experiment 1 were analyzed separately. The data appeared quite similar to Figure 2. This result agrees with a similar analysis of the studies reviewed by Birnbaum (1980). Veit (1978, Experiment 1) used different subjects for the two tasks and also found one rank order for "ratios" and "differences." Note that in this analysis tasks are between subjects, but in each task, standards and comparisons are varied within subjects.

resent predictions of a one-operation theory that will be discussed in the next section (a special case of Equations 3 and 4). Notice that "ratios" and "differences" are nearly monotonically related in all conditions contrary to Figure 3, but the function relating them varies with the modulus and range of examples. These variations in "ratio" judgments can be attributed to different J^* functions in Equation 3 and will be modeled in the next section.

Fit of the theories. Although data for the majority of subjects do not resemble the predictions of Figures 1 and 3 (based on the two-operation theory *and* a power function for subjective heaviness), the two-operation theory might fit better if the assumption of a power function were dropped. Therefore, individual subject data and group data were fit to variations of the two-operation theory and a one-operation theory using the procedure described by Birnbaum (1980), which does not require any assumptions about the psychophysical function. For the one-operation theory, the J and J^* functions were approximated by linear and exponential functions, respectively. Those approximations have been successful for previous research (Birnbaum, 1980). Equations 3 and 4 can then be written as follows:

Table 1
Individual Subject and Group Analyses

	Lack of fit indices					
		Two operations				
Subject	One operation	m = 1.0	m = 1.47	m free	Value of m	Largest "ratio" response
Condition "32"						
1	.022	.097	.067	.029	4.71	32
2	.051	.078	.063	.052	3.28	16
3	.021	.071	.045	.020	4.02	32
4	.051	.123	.092	.054	4.32	32
5[a]	.049	.053	.046	.044	2.16	8
6	.040	.070	.047	.030	4.29	32
7	.034	.071	.048	.029	3.48	32
8	.021	.086	.058	.026	4.56	32
Mean	.018	.069	.043	.019	3.91	25
Condition "800"						
9	.056	.078	.064	.055	2.99	800
10	.028	.048	.034	.028	1.98	1,000
11	.042	.079	.062	.047	3.39	800
12	.063	.086	.072	.063	2.95	1,000
13	.057	.083	.066	.054	3.05	1,500
14	.025	.048	.035	.025	3.35	900
15	.024	.054	.040	.027	3.76	800
16	.029	.047	.036	.031	2.47	800
Mean	.016	.041	.027	.016	3.51	831
Condition "4"						
17	.046	.063	.053	.047	2.28	8
18[a]	.047	.027	.026	.026	1.35	7
19	.045	.058	.049	.045	2.31	5
20[a]	.074	.077	.059	.050	2.79	60
21	.041	.114	.081	.044	4.04	400
22	.064	.088	.076	.065	2.99	4
23	.055	.118	.088	.056	4.03	200
24	.034	.062	.045	.033	2.66	12
Mean	.017	.055	.035	.018	3.53	11

Note. m = an exponent in two-operation theory to represent the inverse of the psychophysical function for numbers.
[a] Data for these subjects appeared to have two rank orders in graphical analyses and were not included in the calculations of the means for their respective conditions.

$$\hat{R}_{ij} = a_R \exp(s_j - t_i) + b_R \quad (6)$$

$$\hat{D}_{ij} = a_D(s_j - t_i) + b_D, \quad (7)$$

where $\hat{R}_{ij}$ and $\hat{D}_{ij}$ are predicted "ratios" and "differences," and a_R, b_R, a_D, b_D, s_j, and t_i are parameters to be estimated from the data. The values of s_1 was fixed to 1.0, leaving 15 scale values to be estimated for each pair of data matrices. Different scale values were estimated for the first and second stimulus to allow for order effects.

To represent the two-operation theory, Equation 6 was replaced with the ratio model

$$\hat{R}_{ij} = a_R(s_j/t_i)^m + b_R. \quad (8)$$

Note that the two-operation theory uses one extra parameter, (m), in addition to those of the one-operation theory. To compare theories with an equal number of parameters, m was initially fixed to a value of 1.47, an exponent presented by Rule and Curtis (1982) to represent the inverse of the psychophysical function for numbers, the output function (judgment function) for magnitude estimation. The value of m is typically observed to be in the range of 1 to 2, and the "ratio" judgments of Rule et al. (1981) are also apparently consistent with a value in this range.

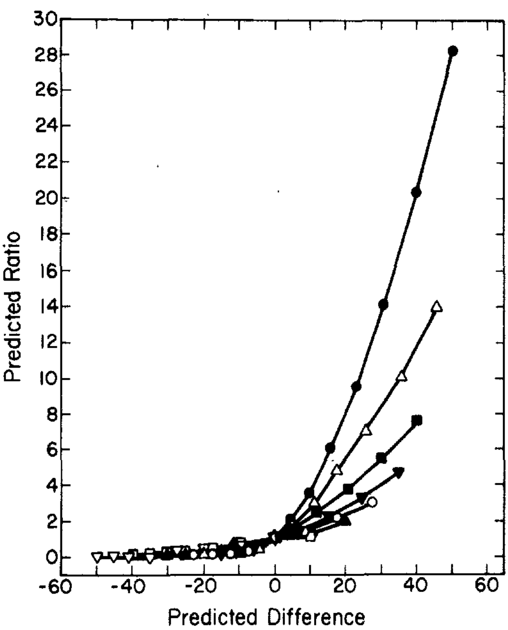

Figure 3. Theoretical ratios plotted against theoretical differences for predictions of Figure 1. (Predicted ratios are subjective ratios raised to the 1.71 power. Symbols correspond to the values of B as in Figure 1.)

The index of fit was the sum of the proportions of total variance in the residuals summed for both "differences" and log "ratios." This index, L, was defined

$$L = \frac{\sum\sum(r_{ij} - \hat{r}_{ij})^2}{\sum\sum(r_{ij} - \bar{r})^2} + \frac{\sum\sum(D_{ij} - \hat{D}_{ij})^2}{\sum\sum(D_{ij} - \bar{D})^2}, \quad (9)$$

where L is the index to be minimized; D_{ij}, $\hat{D}_{ij}$, and $\bar{D}$ are the judged "difference" between stimulus j and i, the predicted "difference," and the mean "difference," respectively. Because the standard errors for the "ratio" tasks vary directly with the mean "ratio," deviations are minimized for the logs of the "ratios" ($r_{ij} = \log R_{ij}$, $\hat{r}_{ij} = \log \hat{R}_{ij}$, $\bar{r}$ = mean $\log \hat{R}$). A computer program that used Chandler's (1969) STEPIT subroutine performed the minimization (Birnbaum, 1980).

Table 1 shows values of L for individual subject and group analyses. Both the one-operation theory and five different versions of the two-operation theory were fit to the data. The value of m was fixed to 1.0, 1.47, 2.0, 3.0, and in one case, it was estimated by the program (m free). The one-operation theory (Equations 6 and 7) gave a better fit to the data than the two-operation theory (Equations 7 and 8) with m = 1.0 for all subjects except Subject 18 (one who did mental arthmetic). When m = 1.47, the one-operation theory fit better than the two-operation theory for all subjects except Subjects 5, 18, and 20.

Birnbaum (1980) noted that as the value of m increases, the two-operation theory approximates the one-operation theory. In other words, as m increases, actual differences approach a weak monotonic function of actual ratios for a fixed design. Thus, large values of m should be taken as evidence favoring the one-operation theory. For 19 of the 24 subjects, fits of the two-operation theory improved as m increased from 1.47 to 3.0. Even when m = 3.0, the one-operation theory still fit better than the two-operation theory for 17 of the 24 subjects. When m was estimated by the program, the best fitting value of m exceeded 2.0 for 22 of 24 subjects, and it exceeded 3.0 for 13 out of 24 subjects.

Separate analyses of group data show that the one-operation theory fit better than or equal to the two-operation theory. Notice that the best fit value of m is greater than 3.5 in all three conditions. To further test the one-operation hypothesis, the theory was also simultaneously fit to the group data for all three conditions. Scale values were assumed to be the same, but the J and J^* functions were allowed to vary across the conditions. For the "difference" tasks, different linear constants were allowed for each condition; in the "ratio" tasks, different linear constants and different multiplicative constants in the exponent were allowed, as follows:

$$R_{ijk} = a_k(\exp[d_k(s_j - t_i)]) + b_k, \quad (10)$$

where R_{ijk} is the ratio of stimulus j to i in condition k. The multiplicative constant in the exponent, d_k, was fixed to 1.0 in one condition but was allowed to vary in the other two. Different constants in the exponent, d_k, mean that responses with different examples will be related by power functions with different exponents.

This version of the theory, which used 28 fewer parameters than the three separate fits of the one-operation theory, fit almost as well as the sum of the three individual fits (.057 vs. .052). Estimated values of s_j were 1.03, 1.40, 1.82, 2.30, 2.80, 3.39, 3.96, and 4.45. Values of t_i were 1.00, 1.39, 1.84, 2.34, 2.87, 3.49, 4.09, 4.60. These estimated values show very small order effects. The linear constants, a_D and b_D, took on values of (2.48 and 0.03), (2.56 and 0.09), and (2.42 and 0.00) for the "4," "32," and "800" conditions, respectively. Values of a_R and b_R were (1.01, 0.00), (0.99, 0.00), and (97.99, 1.37), in the "4," "32," and "800" conditions, respectively. The constants, d_k, were .78, 1.00, and 0.69 in the "4," "32,"

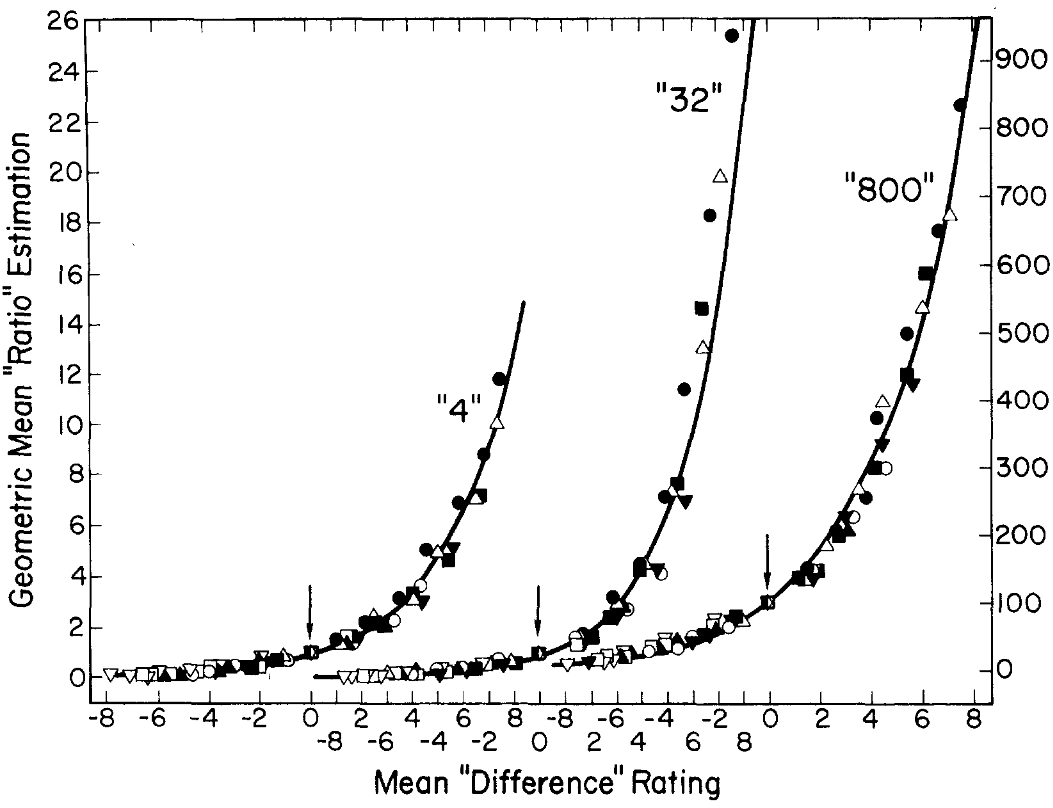

Figure 4. Geometric mean "ratio" estimations plotted against mean "difference" ratings for the three conditions of Experiment 1, as in Figure 3. (Solid lines show predictions of a one-operation theory that assumes different judgment functions for the three conditions but the same comparison operation and scale values for all six data matrices.)

and "800" conditions, respectively. Predictions of this simplified one-operation theory are shown as solid curves in Figure 4. The data points fall close to the predicted curves. Therefore, these results appear consistent with the theory that subjects compute differences of the same scale values in all six matrices, and only the J and J^* functions differ for the different response tasks and different examples.

Experiment 2: Between-Subjects Standards

When standards are varied between subjects. Table 2 presents geometric means for the four conditions of Experiment 2. Data from between-subjects designs can be made to appear either consistent or inconsistent with Equation 1. Notice that when the range of examples is larger for the smaller standard (and a smaller range for the larger standard), the data appear roughly consistent with Equation 1. But when the range of examples is positively confounded with the standards as in Figure 5, the rank order of the data is inconsistent with Equation 1. The data are consistent with the assumption that the J^* functions change for different standards and examples.

Figure 5 shows "ratios" for two conditions of Experiment 2 that violate Equation 1. The curve with solid triangles shows geometric means from the condition with the fourth level of B as the standard and examples that ranged

Table 2
Geometric Means of Between-Subjects "Ratio" Judgments (Experiment 2)

	Level of B (Standard)			
	3 (56.18 g)		4 (84.48 g)	
Stimulus (g)	"4"	"32"	"4"	"32"
20	0.368	0.208	0.220	0.098
35.04	0.498	0.457	0.335	0.217
56.18	0.886	1.027	0.570	0.464
84.48	1.528	1.942	0.994	0.971
120.97	2.319	4.458	1.707	2.022
166.67	3.611	7.585	2.620	4.328
222.66	4.623	12.703	3.374	7.008
290	5.607	17.148	5.073	11.690

Note. Data in Columns 1 and 4 violate the ratio model, whereas data in Columns 2 and 3 are roughly consistent with ratio model predictions (Figure 1). "4" and "32" refer to the example conditions.

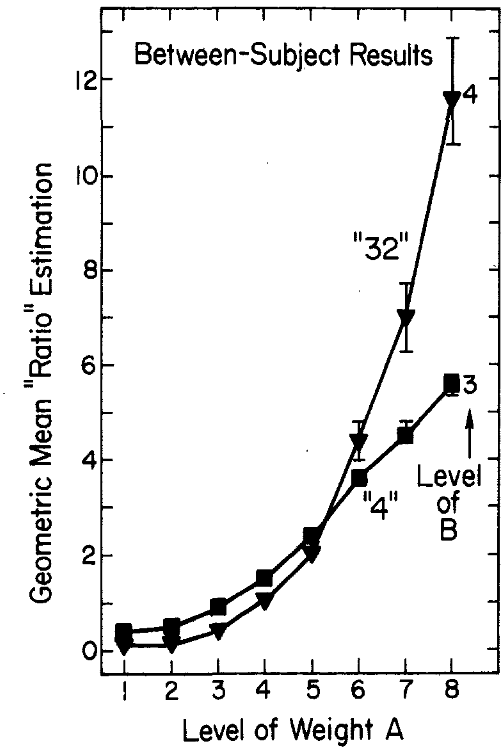

Figure 5. Geometric mean "ratio" estimations for two conditions of Experiment 2, which used a between-subjects design for standards and examples. (The curves connecting solid triangles show means when the examples went up to "32" and the third level of Weight B was the standard. The curves connecting solid squares represent means when the examples went up to "4" and the fourth level of Weight B was the standard. Brackets represent plus and minus one standard error. The crossover interaction is an ordinal violation of the ratio model with the assumption that J^* is the same for both groups.)

to "32": The curve with solid squares shows geometric means from the condition with the third level of B as standard and examples that ranged to "4." Brackets represent plus and minus one standard error. If J^* is the same in the two conditions, Equation 1 implies that variations in the standard between subjects should produce curves that have different slopes but should not cross. The crossover is an ordinal violation of Equation 1.

Two of the between-subjects conditions of Experiment 2 had examples that ranged from ¼ to 4, similar to those of Rule et al. (1981). The rank order of the judgments in those conditions did not replicate the rank order of the corresponding conditions of Rule et al. (1981). For example, in Rule et al. (1981), the judged "ratio" of the 223 g:56 g weight was greater than the judged "ratio" of the 290 g:84 g weight (10 vs. 6). In the present study, the opposite rank order was obtained (4.6 vs. 5). Furthermore, the numerical values of the judgments were much smaller in the present study. For example, in Rule et al. (1981), the judged "ratio" of the 290 g:56 g weight was approximately 11, whereas in the present study, it was only 5.6. The source of the discrepancy is unclear, but it may be due to other uncontrolled factors that influence J^*.

Discussion

Rule et al. (1981) expressed three reservations about the theory that subjects use one operation for "ratios" and "differences": (a) that the conclusion may hold for "metathetic" but not "prothetic" continua, (b) that the stimulus range and spacing may not have been optimally designed in previous tests to detect two operations, and (c) that within-subject variation of standards in the "ratio" task can introduce a bias that could cause "ratios" to be rank ordered the same as "differences."

Prothetic Versus Metathetic

Rule et al. (1981) suggest that the one-operation theory may be appropriate for "metathetic" continua but not for "prothetic" continua. However, this theoretical distinction does not seem to hold up empirically. Not only does there appear to be a single ordering for "metathetic" continua such as pitch (Elmasian & Birnbaum, 1982) and position (Birnbaum & Mellers, 1978) but also for "prothetic" continua such as loudness (Birnbaum & Elmasian, 1977), grayness (Veit, 1978), darkness (Birnbaum, 1982; Mellers & Birnbaum, 1982), and heaviness (Birnbaum & Veit, 1974, and the present results).

Birnbaum (1978, 1982) argued that when the continuum lacks a well-defined zero point, "ratios" are meaningless, and judges will use a subtractive operation. However, for stimuli consisting of *intervals,* both ratio and difference operations can be used (Hagerty & Birnbaum, 1978; Veit, 1978). If visual length is regarded as intervals of position, this distinction may explain the finding of two operations for "ratios" and "differences" of length (Parker, Schneider, & Kanow, 1975).

Stimulus Spacing and Range

Rule and Curtis (1980) noted that when the range of subjective values is small, it is possible that actual ratios and differences will yield similar orders. Therefore, Rule et al. (1981) selected a stimulus range and spacing based on previous scaling to maximize the distinction between "ratio" and "difference" orderings. Experiment 1 used the same stimuli as Rule et al. (1981) but found one order for "ratios" and "differences" in within-subjects designs. The rank order is similar to the rank order of "differences" in Rule et al. (1981). Thus, the finding of one order cannot be attributed to the stimulus range and spacing. Birnbaum and Elmasian (1977, Figure 3) also conducted an analysis of the predicted rank orders and showed that their stimulus levels should have shown two orders given conventional loudness scales.

Within-Subjects Designs for "Ratio" Judgments

Rule et al. (1981), paraphrasing an idea proposed by Stevens and Galanter (1957), state that when the subject receives several standards, "the effective standard on which the subject bases his or her judgment is a compromise between the actual standard and a stimulus near the center of the series" (p. 460). Therefore, they chose to use different groups of subjects for different standards for their "ratio" task. Rule et al. (1981) state that such a bias in "ratio" judgments might explain why "ratios" are monotonically related to "differences." However, the Appendix develops a biased-standard theory from the quotation above and shows that with a factorial design, the orderings of the biased ratios and differences will *necessarily* differ for any positive-valued bias function, if the stimuli are spaced as in Rule et al. (1981). The Appendix also discusses other implications of this biased-standard theory and describes attempts to fit it to the data. Thus, the biased standard notion does not appear to provide an explanation of the present results or previous studies.

A Consistent Account of Both Experiments

An alternative theory that can account for both the present results and those of Rule et al. (1981) is that judges compare stimuli by means of subtraction, regardless of the instructions to judge "ratios" or "differences." However, different groups of subjects have different J^* functions that depend on the stimulus spacing, response scale, range of examples, standard, modulus, and perhaps other features of the experimental procedure. The between-subject "ratios" of Rule et al. (1981) and of Experiment 2 appear consistent with the idea that J^* varies for different groups. This theory merely asserts that different groups of subjects who experience different stimuli use numbers differently to express the same subjective values.

One might argue that if the J^* function is independent of the standard, then it would be possible to compare responses across groups who vary only in standards. Rule (personal communication, 1982) asserts that all subjects in Rule et al. (1981) received the same examples—only the standards changed across groups. If the J^* functions were assumed to be the same, then the question arises of whether to believe results from the within- or between-subjects designs. This question can be stated as follows: Suppose one group of subjects says the "ratio" of A/B is 10, and another group says the "ratio" of A/C is 5. Now suppose that in a different experiment, subjects judge *both* A/B and A/C, and each of those subjects says A/C is greater that A/B. Should we accept the between-subjects (A/B > A/C) or within-subject (A/B < A/C) comparison? If each group of subjects is allowed to have a different J^* function, then the between-subjects comparison can be expressed as $J_1(s_a/s_b) > J_2(s_a/s_c)$, which is not easy to interpret unless $J_1 = J_2$.

Rule (1969) has argued that individual differences in magnitude estimations can be attributed to the output function, J^*. Similarly, Rule and Curtis (1972) allowed the J^* function to vary with different groups of subjects who received different standards. By comparing responses across groups, Rule et al. (1981) implicitly assume that differences in J^* are nonsystematic. However, contextual effects on J^* are systematic and do not "average out." Evidence cited by Poulton (1968), Mellers and Birnbaum (1982), and Mellers (1983) is consistent with the proposition that J^* shows more upward curvature when the standard is in the middle of the stimulus range than when it is at either end.

In summary, if different groups of subjects who receive different standards are allowed different J^* functions, the results of Rule et al. (1981) do not necessarily call for two operations. Rather, the two orderings might be the result of variations in the judgment functions for groups of subjects who received different standards in the "ratio" task. This interpretation can account for both the present results and those of Rule et al. (1981).

Broader Issues

The present results are consistent with the subtractive theory of stimulus comparison. For either "ratios" or "differences," subjects evaluate the algebraic difference between subjective values. The subjective values appear to be independent of the task and response examples. Judgments are a monotonic function of subjective differences, but the numerical properties of the judgment function depend on the response scale ("ratio" or "difference"), the modulus, the examples, the stimulus spacing, the stimulus range, and other between-subjects manipulations of the context.

The subtractive theory has implications for a variety of problems in psychology, including psychophysics, policy analysis, and social judgment. In psychophysics, Stevens (1957, 1971) argues that magnitude estimation is the appropriate method for measuring subjective value. Stevens' power law and related developments were based on the assumptions that the J^* function for magnitude estimation is a similarity transformation and that the comparison operation is ratio. However, the subtractive theory would imply that magnitude estimations are not even a linear function of subjective differences.

The judgment function relating subjective differences to overt responses may be approximated by an exponential function under certain conditions. This exponential function arises from a geometrically spaced set of examples in the instructions (Birnbaum 1978, 1980). In Experiment 1 of the present study, different geometric ranges resulted in responses of "8" or "32" to the same physical ratio (see Table 1). These different ranges in responses could be represented by different powers in the exponential function.

By selecting examples spaced according to other functions, it should be possible to change the form of the J^* function (Mellers & Birnbaum, 1982; Mellers, 1983). Mellers (1983) asked subjects to judge magnitude estimations with no specified examples. In addition, subjects were allowed to select their own standards and moduli. Some investigators (Zwislocki & Goodman, 1980; Zwislocki, 1983) have argued that with this procedure, subjects' responses provide an "absolute" scale of sensation. However, judgments of the same stimuli made in two different stimulus distributions were quite different. Not only single responses but ratios of responses varied with the stimulus spacing. Apparently, "ratio" judgments do not exhibit the invariance properties one would require of actual ratios.

Despite acknowledgment of "biases" in magnitude estimation by many workers in psychophysics (Poulton, 1968; Rule & Curtis, 1982), some investigators use ratio models at face value in studies of policy analysis. For example, Saaty (1977) suggested that ratio models be applied to "ratio" judgments to determine the subjective values of alternatives and the importances of stimulus dimensions. Saaty proposed that these values can be multiplied and summed to evaluate multiattribute alternatives. However, the subtractive theory implies that the values obtained in Saaty's procedure will be exponentially related (with different exponents) to subjective values. To justify the use of a weighted sum, as in Saaty (1977), it should be demonstrated that "ratios" and "differences" of importance are governed by two appropriate operations, and scale values are therefore known to a ratio scale. Competing suggestions for the measurement of importance are given in Birnbaum and Stegner (1981).

In the area of social judgment, ratio models have been proposed to represent equity. Equity is presumed to occur when outcomes (rewards) are in some sense proportional to inputs (contributions). However, when appropriate tests are made, equity judgments show the same contextual effects and violations of ratio invariance as psychophysical judgments (Mellers, 1982; Mellers & Birnbaum, 1982). Mellers (1982) found that inequity judgments are consistent with a subtractive model of stimulus comparison, applied to the relative positions of inputs and outcomes in the two distributions, respectively.

Conclusion

The present experiment investigated the apparent contradiction between the results of Birnbaum and Veit (1974), who found a single rank ordering for "ratios" and "differences" of heaviness, and Rule et al. (1981), who obtained two orderings. Results provide a fairly clear answer to the apparent conflict in the studies and suggest that there is no real empirical disagreement.[3]

In Experiment 1 we used the same stimuli and procedure as Rule et al. (1981) but a within-subject design for standards in the "ratio" task. The results for the majority of subjects agreed with the one-operation theory of Birnbaum and Veit (1974). Therefore, the difference in the findings of Birnbaum and Veit (1974) and Rule et al. (1981) is apparently not due to the stimulus dimension, stimulus values, procedure for lifting the weights, the modulus, or the range of "ratio" examples. In Experiment 2 we used the same stimuli in a between-subjects design for standards in the "ratio" task. The results suggest that the J^* functions differ across groups of subjects who receive different standards and/or examples. Therefore, the difference in results can be attributed to the fact that Rule et al. (1981) used different subjects for each standard in their "ratio" task.

To answer the question of whether one or two operations underlie "ratios" and "differences" of heaviness, a within-subject design should be used, that is, each subject should rank order all standard-comparison pairs. In within-subject designs, variations in the range of examples does not appear to affect the rank

[3] A few minor differences in procedure and analysis remain between the present studies and those of Rule et al. (1981) that deserve a brief mention. First, the present study controlled the order in which the stimuli were lifted (A then B), whereas Rule et al. (1981) used different procedures for the "ratio" and "difference" tasks. In the "difference" task of Rule et al. (1981), the left stimulus always weighed less than the right stimulus, and the left stimulus was always lifted first, thereby confounding position, order, and magnitude. Apparently Rule et al. (1981) assumed no order or position effects. Second, absolute rather than algebraic "differences" were judged. Although these issues can be troublesome in theory (Birnbaum, 1981), the fact that the rank order of "differences" in Experiment 1 is similar to that of Rule et al. (1981) suggests that these variables are not important to the present analysis.

order of stimulus pairs. However, in between-subject designs, the rank order of stimulus pairs does change when different subjects have different standards and response examples. In conclusion, the results for heaviness appear consistent with those for most other psychological continua (Birnbaum 1980, 1982) and appear to be best described by a single subtractive operation.

References

Birnbaum, M. H. (1978). Differences and ratios in psychological measurement. In N. J. Castellan & F. Restle (Eds.), *Cognitive Theory* (Vol. 3, pp. 33–74). Hillsdale, NJ: Erlbaum.

Birnbaum, M. H. (1980). A comparison of two theories of "ratio" and "difference" judgments. *Journal of Experimental Psychology: General, 109,* 304–319.

Birnbaum, M. H. (1981). Reason to avoid triangular designs in nonmetric scaling. *Perception & Psychophysics, 29,* 291–293.

Birnbaum, M. H. (1982). Controversies in psychological measurement. In B. Wegner (Ed.), *Social Attitudes and Psychophysical Measurement* (pp. 401–483). Hillsdale, NJ: Erlbaum.

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

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

Birnbaum, M., and Stegner, S. (1981). Measuring the importance of cues in judgment for individuals: Subjective theories of IQ as a function of heredity and environment. *Journal of Experimental Social Psychology, 17,* 159–182.

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

Chandler, J. P. (1969). STEPIT: Finds local minima of a smooth function of several variables. *Behavioral Science, 14,* 81–82.

Ekman, G., & Sjoberg, L. (1966). Scaling. *Annual Review of Psychology, 17,* 451–475.

Elmasian, R., and Birnbaum, M. (1982). *A harmonious note on pitch.* Unpublished manuscript.

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

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

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

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

Mellers, B. A. (1982). Equity judgment: A revision of Aristotelian views. *Journal of Experimental Psychology: General, 111,* 242–270.

Mellers, B. A. (1983). Evidence against "absolute" scaling. *Perception & Psychophysics, 33,* 523–526.

Mellers, B. A., & Birnbaum, M. H. (1982). Loci of contextual effects in judgment. *Journal of Experimental Psychology: Human Perception and Performance, 8,* 582–601.

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

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

Rule, S. J. (1969). Subject differences in exponents from circle size, numerousness, and line length. *Psychonomic Science, 15,* 294–295.

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

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

Rule, S. J., & Curtis, D. W. (1982). Levels of sensory and judgmental processing: Strategy for the evaluation of a model. In B. Wegner (Ed.), *Social attitudes and psychophysical judgment* (pp. 107–122). Hillsdale, NJ: Erlbaum.

Rule, S. J., Curtis, D. W., & Mullin, L. C. (1981). Subjective ratios and differences in perceived heaviness. *Journal of Experimental Psychology: Human Perception and Performance, 7,* 459–466.

Saaty, T. L. (1977). A scaling method for priorites in hierarchical structures. *Journal of Mathematical Psychology, 15,* 234–281.

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

Shepard, R. (1981). Psychological relations and psychophysical scales: On the status of "direct" psychophysical measurement. *Journal of Mathematical Psychology, 24,* 21–57.

Stevens, S. S. (1957). On the psychophysical law. *Psychological Review, 64,* 153–181.

Stevens, S. S. (1971). Issues in psychophysical measurement. *Psychological Review, 78,* 426–450.

Stevens, S. S., & Galanter, E. H. (1957). Ratio scales and category scales for a dozen perceptual continua. *Journal of Experimental Psychology, 54,* 377–411.

Torgerson, W. S. (1961). Distances and ratios in psychological scaling. *Acta Psychologica, 19,* 201–205.

Veit, C. T. (1978). Ratio and subtractive processes in psychophysical judgment. *Journal of Experimental Psychology: General, 107,* 81–107.

Veit, C. T. (1980). Analyzing "ratio" and "difference" judgments: A reply to Rule and Curtis. *Journal of Experimental Psychology: General, 109,* 301–303.

Zwislocki, J. J. (1983). Absolute and other scales: Question of validity. *Perception & Psychophysics, 33,* 593–594.

Zwislocki, J. J., & Goodman, D. A. (1980). Absolute scaling of sensory magnitudes: A validation. *Perception & Psychophysics, 28,* 28–38.

(*Appendix follows on next page*)

Appendix

Biased Standard Theory

In this Appendix, it will be shown that biased standard theory cannot account for the monotonic relationship between "ratios" and "differences" in a factorial design, if the scale values are spaced as in Rule et al. (1981). Biased standard theory asserts that in within-subject designs, the subjective value of the standard stimulus (denominator) is "biased" toward the average stimulus value. Although Rule et al. (1981) did not formulate the biased standard theory explicitly, the idea translates mathematically as follows:

$$D_{ij} = J(s_j - s_i) \quad \text{(A1)}$$

$$R_{ij} = J^*(s_j/t_i) \quad \text{(A2)}$$

where D_{ij} and R_{ij} are judged "differences" and "ratios," J and J^* are strictly monotonic judgment functions, s_j and s_i are subjective values, and $t_i = T(s_i)$ is the biased value of the standard for the "ratio" task. Although Rule et al. (1981) appear to consider the case when t_i is an average of s_i and a constant, the proof below requires only that the bias, T, is a positive-valued function.

Following Rule et al. (1981), suppose that the stimuli have been spaced such that any four successive scale values have the following properties:

$$s_4 - s_3 > s_3 - s_2 > s_2 - s_1 > 0 \quad \text{(A3)}$$

$$1 < s_4/s_3 < s_3/s_2 < s_2/s_1. \quad \text{(A4)}$$

Successive differences between scale values increase as one moves up the scale, and successive ratios decrease. For example, the numbers 2, 4, 7, and 11 satisfy Equations 3 and 4 because $11 - 7 > 7 - 4 > 4 - 2$, but $11/7 < 7/4 < 4/2$.

Rule et al. (1981) attempt to explain why judged "ratios" of successive stimuli might have the opposite order of subjective ratios. That is, "ratio" judgments, R_{ij}, have the order

$$R_{34} > R_{23} > R_{12} \quad \text{(A5)}$$

in apparent contradiction to Equation 4, if a ratio model is assumed. For stimuli satisfying Equations 3 and 4, Equation 1 implies

$$D_{34} > D_{23} \quad \text{(A6)}$$

because $s_4 - s_3 > s_3 - s_2$, and furthermore,

$$D_{21} > D_{32} \quad \text{(A7)}$$

because $s_2 - s_1 < s_3 - s_2$; therefore $s_1 - s_2 > s_2 - s_3$.

Now, for "ratios" and "differences" to be monotonically related, it must be shown that it is possible to find biased standard values, t_1, t_2, t_3, and t_4, such that

$$R_{34} > R_{23} \quad \text{(A8)}$$

and

$$R_{21} > R_{32} \quad \text{(A9)}$$

as in Equations 6 and 7. However, Equations 8 and 9 imply

$$s_4/t_3 > s_3/t_2 \quad \text{(A10)}$$

and

$$s_1/t_2 > s_2/t_3. \quad \text{(A11)}$$

Therefore,

$$s_4/s_3 > t_3/t_2 \quad \text{(A12)}$$

and

$$t_3/t_2 > s_2/s_1. \quad \text{(A13)}$$

Equations 12 and 13 imply

$$s_4/s_3 > s_2/s_1, \quad \text{(A14)}$$

which contradicts the assumption of Equation 4. Therefore, the biased standard theory cannot explain how "ratios" and "differences" can have the same rank order in a factorial design with scale values that satisfy the Rule et al. (1981) spacing. Note that this proof requires no assumptions about the nature of the bias beyond the assumption that scale values and biased standards are positive. It is also important to note that the biased standard theory cannot be well tested in a triangular design. Other reasons to prefer factorial over triangular designs in psychophysical studies are given by Birnbaum (1981). In sum, if "ratios" are monotonically related to "differences" in a factorial design with scale values spaced according to Equations 3 and 4, then the biased standard theory can be rejected.

Biased standard theory has other testable implications. The "ratios" and "differences" in Figure 2 were rescaled to parallelism separately by means of MONANOVA (Kruskal & Carmone, 1969). Row and column marginal means of the rescaled data are estimates of the scale values. According to biased standard theory, denominator scale values in the "ratio" task should systematically differ from the numerator scale values (and from those in the "difference" tasks). However, the four estimates, numerator and denominator of "ratio" and "difference" tasks, were virtually equal. Furthermore, these marginal means were very close to estimated scale values for the one-operation theory. In summary, the biased standard theory cannot explain how "ratios" and "differences" can be monotonically related, nor is there evidence for any bias in the "ratio" standards.

Received May 31, 1983
Revision received September 26, 1983 ■