# Relative frequencies of numerical responses in ratio estimation[1]

JOHN C. BAIRD,[2] CHARLES LEWIS, AND DANIEL ROMER[3]
*DARTMOUTH COLLEGE*

*This study investigated the frequency of different numerical responses in ratio estimation as a function of the numerical categories actually used by Ss. Analyses were conducted for three series of experiments involving magnitude and free-ratio estimation for the attributes of stimulus length, area, and distance. It was found that Ss use certain numerical categories (e.g., multiples of 1, 10, and 100) much more frequently than others (e.g., 37.5) and that the choice of category depends upon the order of magnitude of the response. The statistical implications for ratio scaling are not dramatic, but are worthy of notice.*

One of the interesting pastimes of the scaling theorist is the determination of mathematical models appropriate for empirical data. A controversial example occurs in psychophysics where theorists attempt to develop mathematical assumptions that can properly be made about data obtained by the method of ratio estimation (Mashhour, 1965; Savage, 1966; Stevens, 1966a). When is an "empirical" ratio scale a "true" mathematical ratio scale and not some other variety, such as interval, ordinal, or nominal?

This problem has been touched upon in work by Stevens (1966), Ekman, Hosman, Lindman, Ljungberg, and Åkesson (1968), Rule (1969), and Engen and Ross (1966). These investigators all concerned themselves, in one way or another, with peculiarities in numerical response behavior. Their work includes theoretical observations as well as empirical treatments of individual differences and the effects of the numerical value of the standard in magnitude estimation.

Another approach to this problem is to examine the frequency of different numerical responses given in ratio estimation. The relative frequencies with which Ss employ different numbers in magnitude or free-ratio estimation might help us decide about the continuity of the numerical response scale presumably being drawn upon in psychophysical experiments. Although it is common knowledge that Ss use some numbers, e.g., multiples of 10, more frequently than others, e.g., 37.5, this S bias seems not to have been directly studied. This paper reports some data relevant to this problem.

Specifically, we have done the following. (1) The relative frequencies of numerical responses were obtained for several ratio-scaling experiments, some published, some not. (2) The relationship between these responses and the stimulus values was noted. (3) The relationship between the responses and the empirically derived curve of best fit was noted. (4) An examination was made of the theoretical effect of a restricted number of response categories upon the mean and standard deviation of judgments.

## EXPERIMENTAL SERIES 1

### Experimental Method

Baird, Romer, and Mathias (1969) used the method of magnitude estimation to obtain ratio judgments of the area of two-dimensional patterns. The standard target was always the smallest in the series and was assigned the arbitrary value of 10 units area. Ss were free to assign any other number (including fractions) to represent the judged ratio between comparison patterns of different objective area and the standard. Over the course of two experiments, involving seven target series, a total of 1,340 numerical responses were obtained from 75 Ss. Here we are interested only in the partitioning of these 1,340 responses into numerical categories.

### Data Analysis

For purposes of this analysis, data from both experiments were lumped together and related to the objective stimulus values

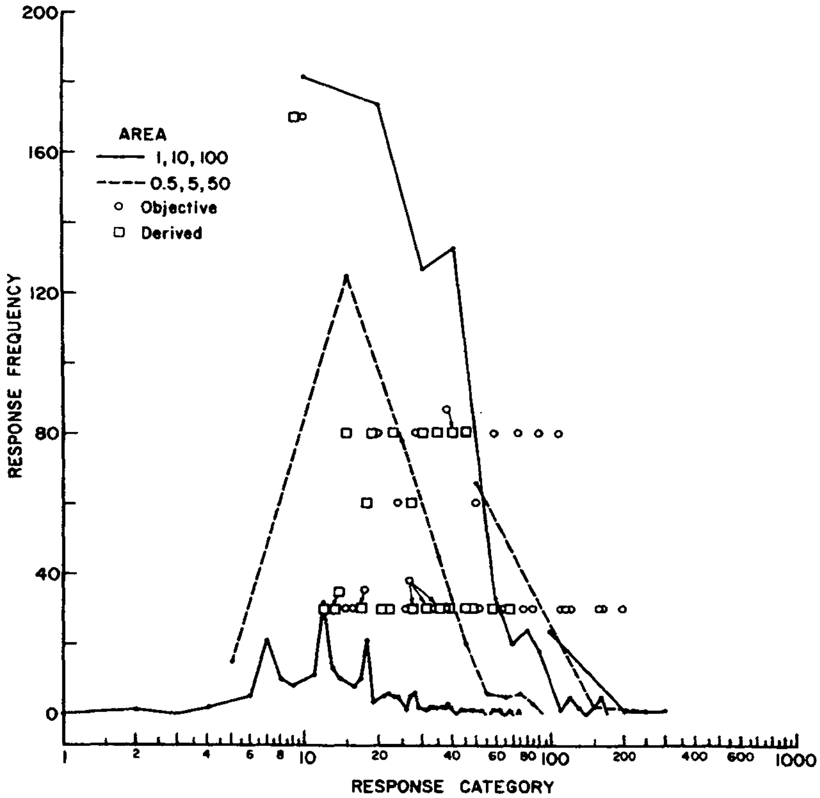


**Fig. 1. Frequency of responses as a function of numerical response category on a log scale. The filled points represent a total of 1,340 magnitude estimates of stimulus area. The solid lines connect values for multiples of 1, 10, and 100, which are grouped independently. The dashed lines perform the same function for multiples of .5, 5, and 50. The open circles represent the expected frequencies for an objective judgment of area. The open squares represent predicted frequencies based upon the best fit of Eq. 1.**

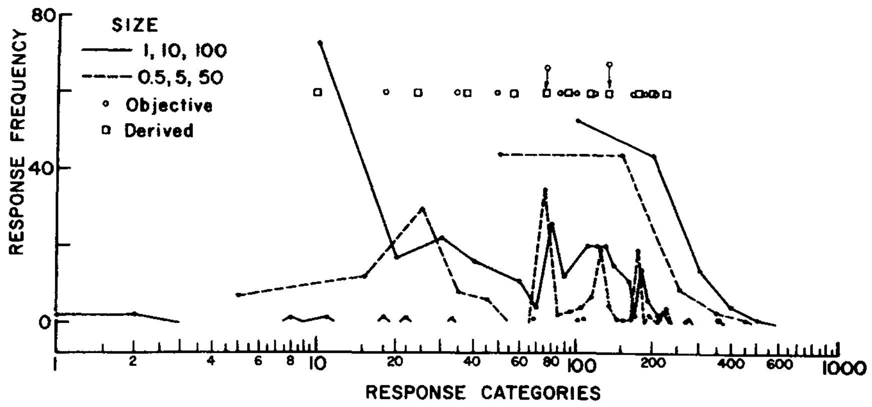


Fig. 2. Frequency of responses as a function of numerical response category. The filled points represent a total of 660 magnitude estimates of line length, triangle area, and the spacing between two dots. Other details are the same as in Fig. 1.

of area by the power function given in Eq. 1,

$$R = \alpha S^{\eta} \qquad (1)$$

where R represents the quantitative value equivalent to the geometric mean of the responses, and S represents the physical area presented. The exponent ($\eta$) and the constant ($\alpha$) were determined by a least-squares fit of a straight line to the logarithmic form of Eq. 1, which is represented by Eq. 2.

$$\log R = \eta \log S + \log \alpha \qquad (2)$$

The value of $\eta$ was .67, and the value of $\alpha$ was 1.97. Next, we computed the response value predicted by the empirical fit of Eq. 1 for each of the stimulus sizes. The hypothetical frequency of each such value was then taken to be the frequency with which the corresponding stimulus was presented. This procedure yields a theoretical response frequency for each stimulus, and this value can be compared with the actual response frequencies generated by Ss. We also compared the theoretical frequency of responses that would be expected if Ss were perfectly accurate in their estimates of physical area. Finally, and most importantly, the actual responses of Ss were tabulated according to the frequency of usage for different numerical categories (e.g., 5, 10, 25, 37.5).

The results of the foregoing analysis are presented in Fig. 1. The frequency of responses is plotted as a function of numerical response categories on a log scale for (1) the Ss' responses (filled circles), (2) the theoretical frequency of responses based upon an accurate (objective) estimate of physical area (open circles, and (3) the theoretical (derived) response frequencies based upon the curve of best fit for Eq. 1 (open squares).

The results in Fig. 1 show clearly that Ss emitted a preponderance of responses that were multiples of 10 and, to a lesser extent, odd multiples of 5. The highest frequency occurs for the number 10, the preassigned value of the standard. These results have been emphasized in Fig. 1 by connecting multiples of 1, 10, and 100 by solid lines and multiples of .5, 5, and 50 by dashed lines. In this regard, it is important to note that all multiples of 1 are connected, but these values are not connected to multiples of 10, which in turn form a group independent of multiples of 100. The same procedure was followed in connecting multiples of .5, 5, and 50.

Examination of Fig. 1 reveals that especially high frequencies do not usually correspond to either the frequencies predicted on the basis of objective area or the frequencies derived from fitting data to Eq. 1. However, the level of the frequencies derived as a function of numerical categories is roughly similar for all three sets of points: objective frequencies, derived frequencies, and subject frequencies. Another observation one can make in Fig. 1 is that a greater variety of numbers seem to be used by Ss at lower values (below about 35) than at higher values.

### Conclusions

Alternative interpretations are possible for these results. The more obvious ones are listed below.

(1) Ss use certain numerical responses more often than others. In the region between 10 and 100, Ss tend to use multiples of 5 and 10, with the spacing between categories increasing with increasing numerical magnitude. (2) Ss emit numbers from 10 to 100 exclusively with the relative frequencies given in Fig. 1. Other multiples of 10 and 5 would not show the same pattern, e.g., the numbers 100 to 1,000. (3) Ss discriminate better around the position of the standard, and, therefore, they use more response categories in that region. (4) The data in Fig. 1 are a unique function of the specific experimental conditions (i.e., area attribute).

Furthermore, the general correspondence between Ss' response frequencies and the theoretical frequencies makes it hazardous to come to any firm conclusions about numerical behavior solely on the grounds provided by this study. For this reason analyses were carried out for several other experiments.

## EXPERIMENTAL SERIES 2

### Experimental Method

This study involved magnitude estimation of lines, triangles, and the spacing between two dots. Judgments of both size and distance were obtained for the same stimuli, which were presented in a dark room with monocular observation through a reduction tube. The experiments actually were conducted several years ago but were never published. Some procedural details are presented in Baird (1970). Each of the three types of pattern was duplicated in 11 different sizes. The middle size served as standard and was assigned the value of 100. Size judgments of the comparison targets were given in respect to this standard, whereas changes in size were treated as changes in target distance when Ss estimated distance. Hence, the same visual-angle changes were stimuli for changes in both size and distance. Thirty Ss were separated into three equal groups, corresponding to the three attributes: lines, triangle area, and dot spacing. Each S made two judgments of size and two judgments of distance for each of the 11 targets. This produced 660 size judgments and 660 distance judgments. Our concern here is with the partitioning of these judgments into different numerical categories.

### Data Analysis

The analysis was similar to that employed for Series 1. The size data were lumped together from the three conditions and treated as though all judgments (including those for triangles) were of linear extent. The geometric means were computed for responses associated with each of the stimuli and results were fit by a straight line in accordance with Eq. 2. The slope (exponent) was 1.25 with $\alpha$ equal to .28. Theoretical responses were calculated from Eq. 1, as well as theoretical values predicted on the basis of the physical lengths of the targets. In each case, the frequency of stimulus presentation was associated with the appropriate response

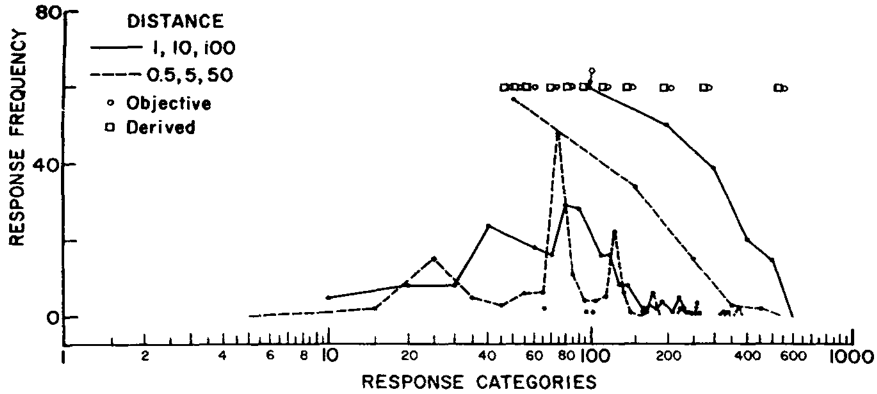


Fig. 3. Frequency of responses as a function of numerical response category. The filled points represent a total of 660 magnitude estimates of distance based upon changes in frontal size. Other details are the same as in Fig. 1.

value. Finally, the actual response frequencies were tabulated for each numerical category used by Ss.

These results are presented in Fig. 2, which is similar in form to Fig. 1. Ss used considerably more numbers that were multiples of 5 and 10 than they did other categories. Above 100, multiples of 50 and 100 seemed predominant. The overall pattern of response frequencies as a function of numerical category does not follow the pattern obtaining for the theoretical and derived frequencies also given in Fig. 2. As in Series 1, the number assigned to the standard received the most responses. Therefore, the use of the number 10 in the first series probably was influenced by the fact that this value was assigned to the standard. It is interesting to note also that the spacing between commonly used categories increased with increases in numerical value.

Distance judgments were expressed as a function of physical distance (derived from geometrical considerations). This transformation of stimulus size to obtain theoretical distance involved taking the reciprocal of each size and multiplying the result by 10,000 in order to reinstitute the distance scale with a standard of 100. Theoretical data were generated on the basis of the derived, best linear fit of Eq. 1. The exponent was 1.001 and the value of $\alpha$ was .94.

The distance data are presented in Fig. 3, which can be interpreted in the same manner as Fig. 2. Ss emitted proportionally more responses that were multiples of 5 and 10 for categories below 100, and primarily multiples of 50 and 100 above. Very few other categories were used, and the responses certainly do not correspond with either the theoretical or derived values shown in Fig. 3. The pattern of frequencies as a function of numerical category seems independent of the theoretical values. This removes some ambiguity in the interpretation of Fig. 1, and offers support for the hypothesis that *specific* response categories are independent of *specific* theoretical and derived values.

### Conclusions

The number of tenable hypotheses posed to explain the results of Series 1 is reduced by the results of Series 2. (1) It appears that Ss definitely use certain numerical categories more frequently than others, and that these responses tend to be multiples of 5 and 10 below 100 and multiples of 50 and 100 above. More generally, the spacing between frequently used categories increases with increases in numerical value.

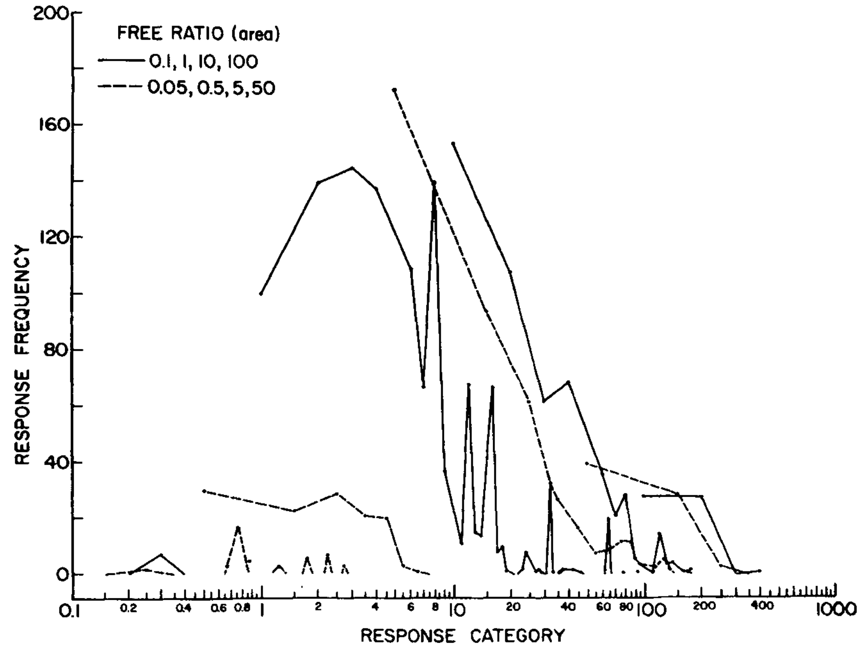


Fig. 4. Frequency of responses as a function of numerical response category. The filled points represent a total of 1,760 free-ratio estimates of stimulus area. Because of the experimental procedure, it is not possible to calculate theoretical or derived frequencies for specific categories. Other details are the same as in Fig. 1.

Series 2 leads us to reject several alternative hypotheses suggested by Series 1. Specifically, (2) the response range over which this phenomenon occurs is not restricted to the numbers 10 to 100, but is present in a modified form for categories between 100 and 1,000. The same pattern of responses occurs below and above the standard. Also, results are not limited to a single standard value, such as 10. (3) An increase in discrimination (as manifested as an increase in the number of response categories) around the standard is not dominant in the results of Series 2. (4) The use of certain categories is not limited to judgments of stimulus area, but also occurs for linear extent, and for distance judgments in a horizontal plane based upon changes in linear, frontal extent.

At least two important questions remain unanswered. First, will the same dominance of categories occur when a standard value is not assigned by the E, that is, when Ss are allowed to choose their own numbers in a free-ratio estimation task? Second, does this phenomenon occur for small numbers, say, between 1 and 10? The final experiment was designed to answer these questions.

## EXPERIMENTAL SERIES 3

### Experimental Method

In this study, Ss were permitted to

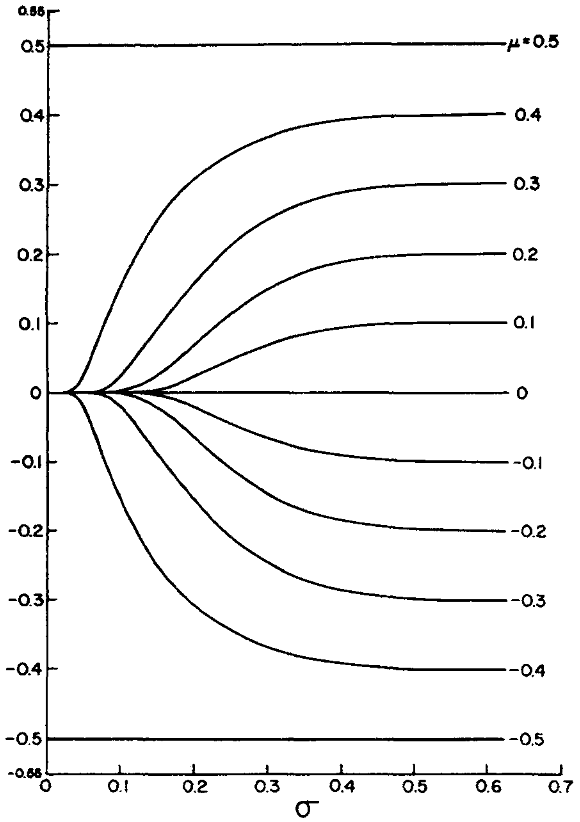


Fig. 5. The effect of rounding (to the nearest integer in this case) on the mean. The mean of rounded observations, $\mu_r$, is given as a function of the true mean and standard deviation, $\mu$ and $\sigma$.

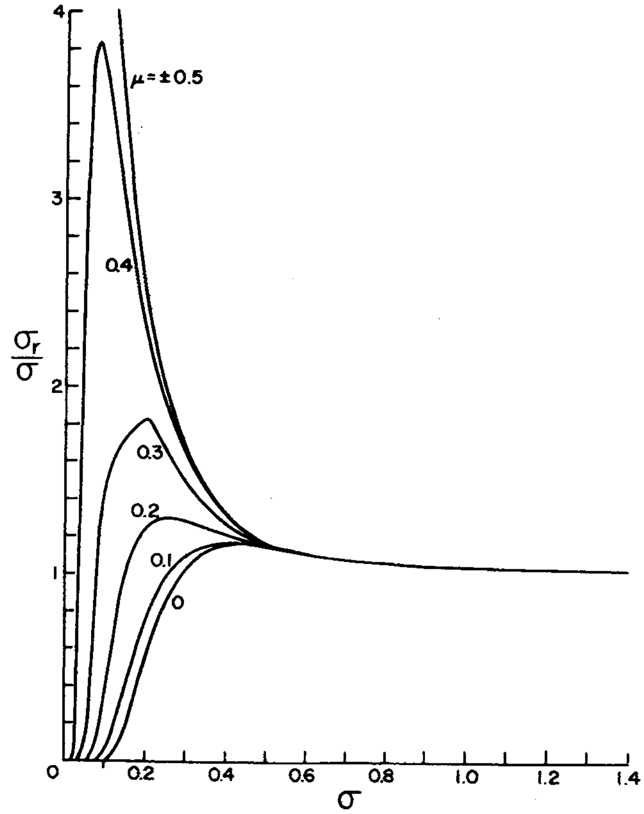


Fig. 6. The effect of rounding (to the nearest integer) on the standard deviation. The ratio of the standard deviation of rounded observations to the true standard deviation, $\sigma_r/\sigma$, is given as a function of the true mean and standard deviation, $\mu$ and $\sigma$.

choose their own numbers to represent the judged ratios between the sizes (areas) of either single squares or patterns comprised of individual smaller squares, generated according to rules described elsewhere (Baird et al, 1969). Six Ss were run on the single squares and five Ss on the complex patterns. The general procedure was similar to that used in Series 1 and 2, except that each S made 20 free-ratio estimates of each of eight targets, which varied in size over approximately one log unit. Patterns were presented randomly in blocks of eight (each shown once), without a standard. Ss were encouraged to use any numbers they wished to express their judgments.

### Data Analysis

The main part of the analysis was similar to those previously described for Series 1 and 2, except that here it is not possible to compute theoretical frequencies on the basis of the objective sizes or on the basis of a derived function. This is due to the fact that a standard was not employed.

Exponents were computed for individual Ss. They varied between .49 and .87, a range comparable to that suggested by Baird (1970) for this type of judgment. These values were based on all 20 judgments. Exponents also were computed separately for the first 2 trials, and for the first 10 trials. These exponents were almost identical to those based upon data from all 20 trials.

The frequency of different numerical categories is plotted in Fig. 4, where interpretation is similar to that for Figs. 1-3. In particular, the pattern of response frequencies is similar to the cases where a standard was employed. That is, multiples of 5 and 10 were used in the range from 10 to 100. There is also a suggestion of the predominance of multiples of 50 and 100 in the range from 100 to 1,000.

A sufficient number of responses were obtained between 1 and 10 to provide a good indication of the categories used in this range. Multiples of 1 predominated, while multiples of .5 were used to a lesser extent.

### Conclusions

These results confirm and extend the generality of Hypothesis 1 that Ss use certain numerical categories more frequently than others, and that the choice of categories is sytematically related to the order of magnitude of the response.

## STATISTICAL IMPLICATIONS

In light of these findings, scaling theory might still hold that Ss think on a continuous scale and round off to convenient numbers when they give a verbal response. Without pressing for the details of this operation, it is possible to study the bias such a procedure would introduce.

For convenience, assume that the value which is rounded when the response is made comes from a normally distributed population with mean $\mu$ and standard deviation $\sigma$. The response will then have a mean $\mu_r$ and standard deviation $\sigma_r$, different from $\mu$ and $\sigma$. Mathematically, these two parameters can be described by the following equations:

$$\mu_r = \sum_{i=-\infty}^{+\infty} [ri \int_{r(i-.5)}^{r(i+.5)} \varphi(x)dx] \quad (3)$$

$$\sigma_r^2 = \sum_{i=-\infty}^{+\infty} [(ri - \mu_r)^2 \int_{r(i-.5)}^{r(i+.5)} \varphi(x)dx], \quad (4)$$

where $\varphi(x)$ stands for the normal density function, with mean $\mu$ and standard deviation $\sigma$. In these equations, the rounded responses are represented by ri, so that possible responses would include 0, ±r, ±2r, ±3r, with a basic rounding unit r.

With the help of an electronic computer and numerical integration, Eqs. 3 and 4 were evaluated for different choices of $\mu$ and $\sigma$. The results of these calculations are shown in Figs. 5 and 6. The rounding unit was arbitrarily chosen to be unity, and $\mu$ was varied in steps of .1 from −.5 to +.5. Changing $\mu$ by any integral number of rounding units would, of course, leave the results unchanged. Similarly, changing the scale of all values would be sufficient to give the results for any other rounding unit.

Figure 5 shows that, when $\sigma$ is small enough, $\mu_r$ may be biased by almost half a rounding unit. However, for any choice of $\sigma$ greater than half a unit, the bias of $\mu_r$ becomes negligible, regardless of the value of $\mu$.

The picture for $\sigma_r$ is slightly more complicated, as indicated by Fig. 6. When $\mu$ = .5 and $\sigma$ is sufficiently small, $\sigma_r$ is a constant (.5). This is because half the observations will be rounded to 0 and half to 1. As a result, the ratio $\sigma_r/\sigma$ increases without bound as $\sigma$ approaches 0. For other values of $\mu$, $\sigma_r/\sigma$ eventually approaches 0 as virtually all observations are rounded to the unit nearest $\mu$.

When $\sigma$ is greater than .6, the position of $\mu$ has little influence on $\sigma_r$, but the common value of the $\sigma_r$ continues to be somewhat greater than that of $\sigma$ as the latter increases beyond the size of the rounding unit.

In summary, the statistical effects of the observed behavior are not dramatic. Nonetheless, it may be worthwhile for future experiments to compare the apparent rounding interval employed by the Ss at any given level with the observed standard deviation of the responses at that level. This information could be employed to evaluate the maximum bias that might be expected to appear in the responses.

In practice, this comparison is complicated by the fact that the standard deviation is dependent on both the stimulus magnitude and the choice of the standard (Baird, 1970). Moreover, we have shown here that the rounding interval varies as a function of stimulus magnitude. Consequently, a more comprehensive treatment of this problem is desirable and is presently being considered.

## REFERENCES


BAIRD, J. C. *Psychophysical analysis of visual space*. London: Pergamon Press, 1970.

BAIRD, J. C., ROMER, D., & MATHIAS, C. G. Area and randomness estimates of complex stimuli. Perceptual & Motor Skills, 1969, 28, 659-669.

EKMAN, G., HOSMAN, B., LINDMAN, R., LJUNGBERG, L., & Åkesson, C. A. Interindividual differences in scaling performance. Perceptual & Motor Skills, 1968, 26, 815-823.

ENGEN, T., & ROSS, B. M. Effect of reference number on magnitude estimation. Perception & Psychophysics, 1966, 1, 74-76.

MASHHOUR, M. A note on the validity of the power law. Scandinavian Journal of Psychology, 1965, 6, 220-224.

RULE, S. J. Subject difference in exponents from circle size, numerousness, and line length. Psychonomic Science, 1969, 15, 284-285.

SAVAGE, C. W. Introspectionist and behaviorist interpretations of ratio scales of perceptual magnitudes. Psychological Monographs, 1966, 80, No. 19, Part I, 1-32.

STEVENS, S. S. Operations or words? Psychological Monographs, 1966a, 80, No. 19, Part II, 33-38.

STEVENS, S. S. On the operation known as judgment. American Scientist, 1966b, 54, 385-401.


## NOTES


1. This research was supported in part by a grant from the U.S. Public Health Service (5 R01 MH 14379-02).



2. Address: Department of Psychology, Dartmouth College, Hanover, New Hampshire.

3. Current address: Department of Psychology, University of Illinois at Chicago, Chicago, Illinois 00000.



*(Accepted for publication February 9, 1970.)*