

Psychophysical Study of Numbers

I. Generation of Numerical Responses *

John C. Baird and Elliot Noma**

Dartmouth College, Hanover, New Hampshire

Received April 30/December 31, 1974

Summary. Two experiments were conducted to study the number biases of subjects in situations not involving the usual psychophysical stimuli. In Exp. I subjects were asked to generate numbers (within boundary conditions) they thought other people would produce under the same conditions. In Exp. II only a single lower boundary (e.g., 1, 10 or 100) was employed and subjects generated a set of numbers larger than the boundary. Results suggested that definite number biases exist. Multiples of 1, 10, 100 and to a lesser extent 5, 50 and 500 dominate and are appropriate to the log cycle. That is, multiples of 1 occur most often in the cycle 1—10, multiples of 10 in the cycle 10—100, etc. The implications of these results are noted for several psychophysical theories.

Francis Galton (1880) once noted that several of his intellectual acquaintances had images of numerals located in one-, two-, or three-dimensional visual space. The special arrangement of the numerals was fixed for an individual but varied widely among people who possessed this ability (estimated by Galton to occur in 1 of 30 males and in 1 of 15 females). The sample conceptions he reports are quite bizarre, but some imply a logarithmic relation between imaged numerical spacing and arithmetic spacing while others reveal the importance of pivotal numbers such as 1, 10, and 100. However, no clear invariances emerge from the conceptions of these unique individuals.

The impetus for our own interest in this topic comes from modern psychophysics, where numbers are used extensively as response categories. The present paper is the first in a fourpart series documenting people's conception and use of numbers in a variety of contexts most closely associated with the field of psychophysics, although we feel the results also have implications for the general field of cognitive psychology. In the course of this series, empirical data and theoretical models are presented to elucidate the transformation between physical and perceptual scales, as well as the similarities between the processing of numbers and the processing of more usual physical attributes such as sound, light, and length. The final paper in the series outlines a new theory of scaling (Preferred State Theory), which applies to a variety of physical attributes and which is based on a generalization of the principles assumed to underly perception of the number continuum.

Recent approaches to psychophysical scaling (e.g., magnitude estimation), in contradistinction to Galton's data, tacitly assume *Ss* have at their disposal a

* We thank Charles Lewis of the University of Illinois for advice on statistical problems.

** Presently at the University of Michigan.

perceptual scale of equally spaced numbers in one-to-one correspondence with values along the physical number scale (Stevens, 1966). If one believes there is a direct link between the perception and use¹ of numbers, the type of attribute scale (to which numbers are matched) can be derived. It, therefore, is quite important that the perception of numbers be thoroughly explored empirically in order to provide a solid foundation for theoretical assumptions.

The usual attack on this problem involves the "scaling" of numbers after the fashion employed for any other attribute. The underlying purpose is to determine whether numbers are subject to the same perceptual laws as other stimuli, and if they are, to then derive the psychophysical function from knowledge of the number scale. However, here, as in all of psychophysics, the method has a hand in dictating the outcome, and different methods lead to different relationships between perceptual and physical numbers. This, of course, leads to theoretical difficulties.

The simple view that perceptual number is a power function of physical number with an exponent of 1.0 does receive support from experiments by Rosner (1965) who had *Ss* match numbers from one region of the scale with numbers from another. On the other hand, Attneave (1962) originally speculated that the power function for number had an exponent of about .4, and this notion gains support from recent work (Curtis, Attneave and, Harrington, 1968; Curtis, 1970). In these experiments, the size of the relevant "output" or "number" exponent was not always the same, indicating perhaps that the exact form of the number scale is not independent of the second attribute under study.

Contrary to these lines of reasoning, Ekman (1964) proposed that perceptual number is a log function of physical number, and he offered data to back this claim (Ekman and Hosman, 1965). If this view were correct, the power function would result from the combination of two log functions, one for number and one for ordinary physical stimuli. Rule (1969) comes to similar conclusions based on a Thurstonian approach.

Unfortunately, there are subtle, but fundamental, drawbacks with all the research outlined or implied in the above review. First, if we are unsure about how people use numbers, it seems unwise to accept uncritically some convenient answer (such as one-to-one correspondence) to serve as a standard (McGill, 1960). Second, standards for stimulus attributes do not exist by which scales for the perception of number can be obtained. In other words, psychophysical methods are inappropriate for obtaining scale values for numbers because in previous studies, these values were not independent of other attribute scales whose characteristics were themselves uncertain. With available techniques, we cannot scale one attribute without somehow involving another.

The obvious alternative is to secure estimates of scale values by techniques other than those used in psychophysics. And in fact, it has been possible to do this by looking at *Ss*' preferences for numbers. Most such studies have used the integers one through nine. Preferences were expressed in a variety of ways, and response frequencies were tabulated from large groups of *Ss* (Eagleston and Lipford, 1944; Heywood, 1972; Ross and Kohl, 1948; Winick, 1962; and Yule, 1927). Each of these investigators offered a psychological interpretation of the rank order

¹ That is, the perceptual transformation is revealed by the behavioral use of numbers.

of integer preferences found in their work. Unfortunately, each study uncovered a *different* rank order. To discover possible consistencies across these experiments, we computed rank order correlations on the major rankings reported in the five studies. Of the 10 possible pairings ($5 \times 4/2$), 7 correlations were negative, 3 positive, and none significant ($df = 7$, $p < .05$). The conclusion seems justified that there are no preferences among the integers one through nine which transcend a range of experimental tasks. The total response frequencies from all studies can be seen as distributed in a rectilinear fashion across these integers.

However, this effect is limited to the integers 1 to 10 since data collected by Banks (1973) and by Banks and Hill (1974) indicate that in order to maintain equal spacing along the perceptual scale, spacing between physical numbers must be successively increased for magnitudes greater than 10. These investigators asked *Ss* to randomly generate a series of numbers which *E* could then rank order. The theoretical rationale for this technique is straightforward. Assuming a perceptual scale exists and that the spacing between values along that scale are equal (Fechner's assumption, 1907), random sampling² should yield appropriate measures in terms of the corresponding physical number continuum. By then placing the data in an increasing order of magnitude, the relationship between perceptual (equal-spaced ranks) and physical (generated) values can be assessed by standard curve-fitting procedures (Banks, 1973; Banks and Hill, 1974). As determined by this technique, these authors found that either a log or power function (with an exponent of about .66) fit the relation between perceptual and physical magnitudes. In a further series of experiments, the applicability of this approach was demonstrated for several other attributes (e.g., length, sound, duration). Most importantly, the resulting functions bear a close resemblance to those obtained by other methods, such as magnitude estimation, inasmuch as the exponents obtained from the two approaches are a reasonably constant proportion of each other.

On the other hand, response frequencies from actual psychophysical experiments indicate that the perceptual scale is linearly related to the physical scale within a log cycle (e.g., 1 to 10, 10 to 100, 100 to 1000) but logarithmic across cycles (Baird, Lewis, and Romer, 1970). This type of result could lead to the erroneous impression of a log or power function for cases where a large range of numbers was involved and where details of the function were ignored. This analysis also showed that *Ss* "preferred" to use certain numbers much more frequently than others. Specifically, multiples of 1, 10, and 100 were produced quite often, together with multiples of 5, 50, and 500. These results agree with those on number preferences reported earlier by Ross and Engen (1959). They had *Ss* guess numbers within a range of 1 to 120, where guesses continued until the correct (predesignated) number was hit upon. Favorite guesses tended to end in 0 or 5; only a small percentage of numbers ended in the other eight digits.

In the present experimental series, we in effect combined the approaches of Ross and Engen (1959) and Banks (1973) into a single procedure, although at the inception of the study, we were unaware of their work. In one experiment, *Ss* were told to generate numbers (within boundary conditions) which they thought other *Ss* would generate under the same circumstances. In a second experiment, they

² It is apparently quite difficult for *Ss* to generate a random sequence of items (Wagenaar, 1972), but this should not seriously affect our conclusions.

were permitted to generate any number which was larger than a single lower boundary. No standard psychophysical stimuli were used. We were primarily interested in determining what types of numbers would be produced in such tasks. Because such results have implications for a variety of theoretical positions, some of which are developed in the later papers of this series, a variety of analyses were undertaken. In particular, we looked at frequencies of numerical responses within different ranges, uncertainty measures which have relevance for an earlier psychophysical theory (Baird, 1970 a, 1970 b, 1970 c), as well as general functions between perceptual and physical scale values (after Banks and Hill, 1974).

Experiment I

The chief purpose of Exp. I was to obtain information about number biases without involving usual psychophysical stimuli. The *Ss* were required to generate numbers (within numerical boundaries) which they thought other people would give under the same circumstances. This procedure was employed because pilot work indicated that some *Ss* approached the task as a challenge to their ingenuity in producing unusual numbers. By asking *Ss* to predict other people's responses, we forced them to reveal personal number biases, since we assume *Ss* had no independent evidence for assessing the bias of others in this task. Results were compared with those found in psychophysical studies (Baird *et al.*, 1970).

Subjects. The participants were 60 male undergraduates enrolled in an introductory psychology course at Dartmouth College.

Stimulus Materials. The stimulus ranges were typed at the top of individual sheets of paper which were stapled together in a different random order for each *S*. Ten spaces were marked on each sheet to indicate that *S* was required to write down 10 numbers for each stimulus range. Fifteen ranges were used: 1 to 10, 1 to 100, 1 to 1000, 10 to 100, 10 to 1000, 100 to 1000, 11 to 99, 5 to 95, 12 to 989, 13 to 85, 14 to 782, 17 to 100, 1.3 to 9, 1.9 to 942, and 15 to 985. Henceforth, the first six ranges will be referred to as *unit-digit*, and the remaining nine ranges will be called *multiple-digit*. These particular cases were chosen to secure a representation of ranges covering more than one log cycle, and to gain variety in the boundary numerals.

Procedure. The experiment took place in a small auditorium which accommodated all *Ss* simultaneously. Instructions noted that a large group of people had already participated in the study and that *Ss*' task was the same as theirs: to write down 10 numbers in the available spaces for each stimulus range. Subjects were instructed to produce responses that reflect a prediction of the numbers most likely generated by the other people. In order to simulate the instructions given in an experiment employing magnitude estimation, *Ss* were told they could use fractions as well as whole numbers and that the boundary numbers were permissible responses. In addition, they were allowed to repeat any of their responses for different ranges. The *Ss* recorded responses at their own pace.

Results and Discussion

The group mean and standard deviation of all responses (across all *Ss*) were calculated for each stimulus range³. The relationship between these two elementary statistics can tell us a great deal about *Ss*' judgment strategies. For example, one could suppose that *Ss* conceive of the number scale with the same perceptual spacing between each pair of numbers at all locations. This is the implicit assumption made by most investigators using the method of magnitude estimation (Stevens, 1966). Under this constant-interval model, increasing the stimulus range (e.g., 1—10 to 10—100) should add a constant to the mean but should leave the

³ We also analyzed several transformations of the data. For example, the frequencies of numbers were weighted by taking logarithms or inverses of their rank order. The use of these transformed values did not alter the results substantially.

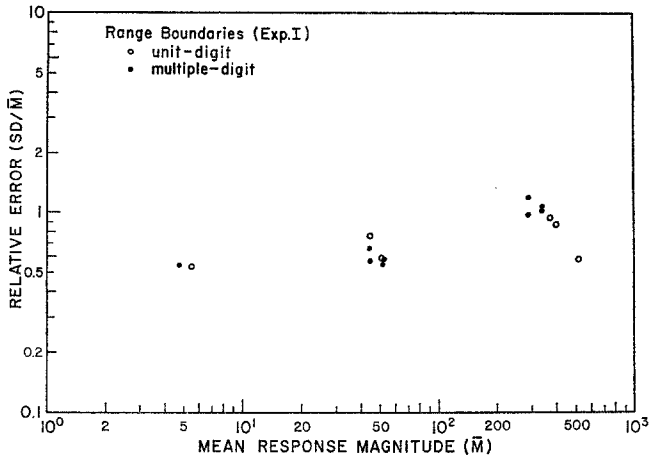


Fig. 1. Relative error (standard deviation divided by the mean) as a function of mean response magnitude. The coordinates are logarithmic. Points are based on the group data obtained in Exp. I for unit-digit (open circles) and multiple-digit (closed circles) ranges

standard deviation unaffected. Therefore, the ratio of the standard deviation to the mean (relative error) would decrease linearly as stimulus range increased. Alternatively, it can be proposed that *Ss* maintain a constant ratio among numbers generated between stimulus ranges. When moving from one range to another, it is as though *Ss* multiply each number (e.g., 1, 2, . . . 9) in the first range by a constant in order to generate numbers in the second range. If this strategy were followed, both the mean and the standard deviation would change by the same multiplier. Hence, the relative error would be constant for all ranges. These two models (constant interval and constant ratio) are the most likely, but of course, *Ss* may also use some combination rule.

The relevant data for each of the stimulus ranges are presented in Fig. 1 where relative error (SD/\bar{M}) is plotted as a function of the mean in logarithmic coordinates. The open circles represent unit-digit ranges; the filled circles represent multiple-digit ranges.

On the log-log plot, the constant-interval model would be upheld by a linear function with a slope of -1 but with an arbitrary y-intercept. The constant-ratio model would be represented as a horizontal line with a slope of 0 and an arbitrary y-intercept. The data are in closest agreement with the constant-ratio model, although the points do not fall on a perfectly horizontal line (not drawn on the graph). Rather, the function appears to rise somewhat with an increasing mean. This upward trend is less conspicuous for unit-digit ranges. In this regard, it is interesting to note that the position of the mean in respect to the entire log cycle (1 to 10, 10 to 100, 100 to 1000) decreases slightly as the boundaries increase (read along the x-axis). Although not clear from Fig. 1, the standard deviations also increase slightly faster than would be predicted if only a constant multiplier had been applied. The combination of these two facts creates the slight rise in the function. For one-cycle, unit-digit ranges (1 to 10, 10 to 100, 100 to 1000), the

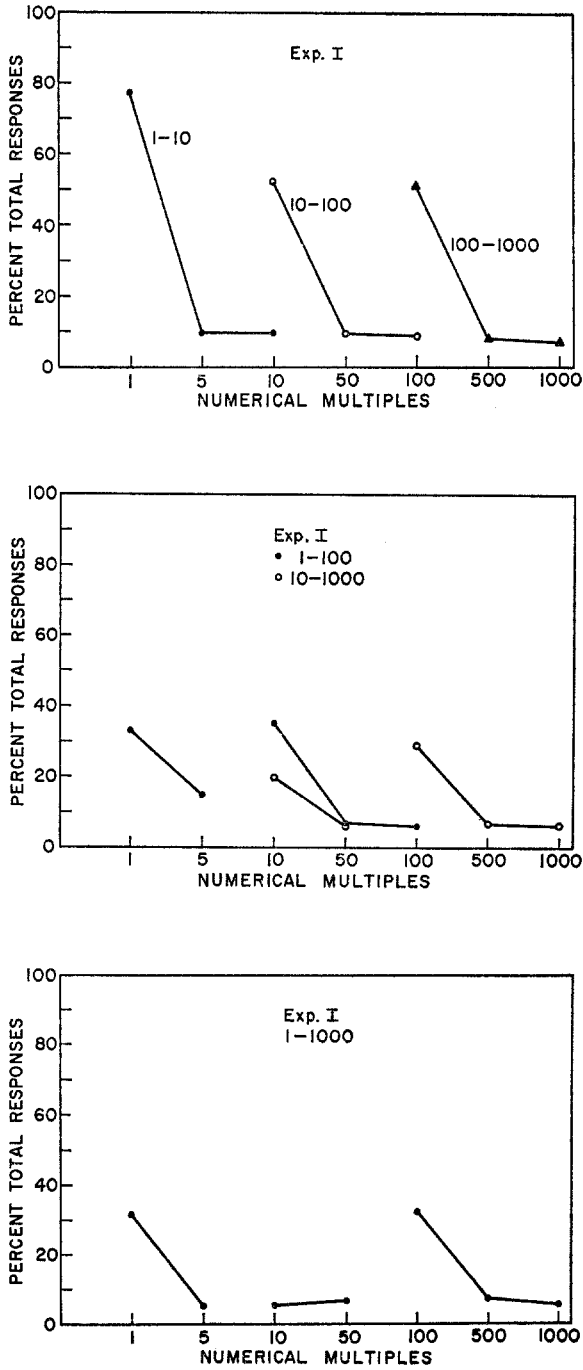


Fig. 2. Percent of total responses (group-data) which are multiples of the categories given on the x-axis. Inclusion in a response category means the response was not a multiple of a higher category listed on the x-axis. Data are given for single-cycle log ranges (top), two-cycle log ranges (middle), and the three-cycle range (bottom)

function appears relatively flat. These three points are shown in Fig. 1 as the lowest open circles within each of the three major clusters of points. This picture is not found for multiple-digit ranges. In general, however, results suggest that *Ss* have a basic conception of the number scale within a standard log cycle and that they act as though changes in the magnitude of the cycle are accommodated by applying the appropriate multiplier to each number in the standard.

More detailed evidence for this conclusion is obtained from the specific response categories generated with ranges of different magnitude and location. It was evident that *Ss* used a preponderance of numbers which were multiples of 1, 10, 100, and, to a lesser extent, multiples of 5, 50, 500. This was especially true for unit-digit ranges. This effect is clear from Fig. 2. There we have plotted the percent of total responses (from the group) which are multiples of the integers shown on the x-axis. In order to qualify for inclusion in a category, a number had to be a multiple of that category but not a multiple of any higher category given along the x-axis. For example, the number 25 is a multiple of both 1 and 5. It would be included in the 5 category but not in the 1 category. The pattern is similar, though depressed, for the remaining ranges since more significant digits occur in the responses.

Fig. 2 (top) shows the results of this analysis for the three single-cycle ranges. The pattern is pronounced and similar for each. A very high percentage of responses are multiples of the lower stimulus boundary, whereas multiples of 5, 50, or 500 are much less frequent and occur about as often as multiples of the upper boundary (the only number possible for this category). Overall, approximately 95% of the total responses is represented in Fig. 2 (top) for the range 1 to 10, approximately 70% for the range 10 to 100, and approximately 65% for the range 100 to 1000. The large percentage obtained for the 1 to 10 range is due to the fact that *Ss* do not give fractions, whereas more significant digits are used in the higher ranges.

The results in Fig. 2 (middle) are for the ranges containing two log cycles, two cases (1 to 100 and 10 to 1000). The pattern is identical for both. The lines connecting the data have been interrupted to separate log cycles and thereby allow easier comparison with the results for single cycles (Fig. 2, top). As was true for one-cycle ranges, the higher range (10 to 1000) has a lower percent of the total responses attributable to the multiples given in Fig. 2, but in both instances, the picture is represented well by a few key response categories.

The results for the range 1 to 1000 are given in Fig. 2 (bottom). The situation resembles that for shorter ranges, except that multiples of 10 are diminished. This probably is not a random effect since the same depression occurs for the range 1.9 to 942. Whether the effect is reliable or not, we cannot explain it. In any event, something on the order of 90% of all responses is represented in Fig. 2 (bottom).

In summary, these results lend weight to the hypothesis that *Ss* use a constant-interval model of the number scale *within* a log cycle and a constant-ratio model *across* cycles. This hypothesis is further supported by the results of previous studies of number preference discussed earlier, and by psychophysical data (Baird *et al.*, 1970). The latter agreement suggests that our experimental procedure taps the same perceptual scale used by *Ss* in judging the intensities of psychophysical stimuli. One must add, finally, that these conclusions must be tempered when

Table 1. Kolmogorov-Smirnov tests (Exp. I). Table entries indicate nonsignificance. Asterisks are used for the log cycles from unit-digit ranges; crosses are for multiple-digit ranges

		Range and Cycle																									
		1A	1B	1C	2A	3A	3B	4B	5B	5C	6C	7B	8A	8B	9B	9C	10B	11B	11C	12B	13A	14A	14B	14C	15B	15C	
Range and Cycle	1A	-	*			*																					
	1B		-			*			*																		
	1C			-	*		*		*	*		+			+								+	+			+
	2A				-	*				*	*																
	3A					-			*																		
	3B						-	*		*	*		+		+												
	4B							-	*	*	*												+				
	5B								-								+							+			+
	5C									-	*		+		+												
	6C											-															
	7B												-						+								
	8A													-													
	8B														-												
	9B															-				+							+
	9C																-				+					+	+
10B																	-										
11B																		-								+	
11C																			-						+	+	
12B																				-				+		+	
13A																					-						
14A																						-					
14B																							-				
14C																								-		+	
15B																									-	-	
15C																										-	

Code	Range	Code	Range	
1	1-1000	8	5 - 95	Cycle A = 1 to 10
2	1- 10	9	12 -989	Cycle B = 10 to 100
3	1- 100	10	13 - 85	Cycle C = 100 to 1000
4	10- 100	11	14 -782	
5	10-1000	12	17 -100	* not significant
6	100-1000	13	1.3- 9	+ not significant
7	11- 99	14	1.9-942	p > .05, two-tailed test
		15	15 -985	

applied to stimulus ranges with boundary stimuli containing more than significant digit.

Kolmogorov-Smirnov Tests. A more detailed analysis of the similarities among the response distributions was obtained by testing for differences among log cycles both within and between stimulus ranges. A Kolmogorov-Smirnov two-sample test was used for this purpose. The entire response distribution for each range was separated into one, two, or three log₁₀ cycles (as appropriate), and each cycle was labeled according to its magnitude (A = 1 to 10, B = 10 to 100, C = 100 to 1000). This coding leads to 25 cycles which, when compared pairwise, create 300 applications of the test. Several further decisions had to be made in order to apply this test. First, the number range within each cycle was split into 100 equal intervals. Second, the multiple-digit ranges were treated further by ignoring the boundary numbers in all comparisons. This was done to eliminate the effect of the standards and was carried out for unit-digit ranges whenever these cases were compared with cycles from multiple-digit ranges. In short, comparisons between two distributions always involved exactly the same section of the scale. For example, in comparing the cycle 1.3 to 9 with the cycle 1 to 10, the relevant distribution in both cases would be 2 to 8. We felt this procedure would permit a fairer test for differences between distributions.

The results are presented in Table 1. Each of the 15 ranges has been labeled together with the appropriate cycle under consideration. The entries indicate that

no significant difference was found between the pair of distributions ($p > 0.05$, two-tailed test). The asterisk represents comparisons among cycles containing one significant digit (unit-digit) in each boundary number. The cross is used for all other comparisons.

An entry either of an asterisk or a cross in Table 1 indicates that no significant difference exists between the two relevant distributions which, in turn, can be used in favor of the hypothesis that *Ss* have a basic conception of the number scale within a log cycle and that they behave as if a multiplier is used in moving from one cycle to another. In this regard, the results are particularly illuminating in showing a distinction between unit-digit ranges and multiple-digit ranges. The former results support the hypothesis; the latter do not. Specifically, it is helpful to separate Table 1 into three sections and discuss each in turn. The three sections are (1) cases where *both* ranges have unit-digit boundaries (coded 1 through 6), (2) cases where both ranges have multiple-digit boundaries (coded 7 through 15), and (3) mixed cases; that is, where one range is a unit digit (1 to 6) and the other is a multiple digit (7 to 15).

Referring to Table 1, there are 10 separate log cycles (1A through 6C) for the pure unit-digit case. Hence, there are $(10 \times 9)/2 = 45$ possible comparisons, shown in the left-most part of the Table. The asterisks indicate that 18 of these are not significant. A rather surprising pattern can be extracted from these findings. Namely, most asterisks (13) occur for comparisons involving the highest cycles in the two ranges under consideration, and three more nonsignificant instances occur for comparisons involving the second highest cycles. In all, 16 out of 18 asterisks follow this pattern, including all but two of the available possibilities. Apparently, the matching of cycles occurs between the highest cycles in the two ranges. This means that the specific cycle (A, B, or C) is *not* treated identically by *S* across ranges. The important thing for a match (nonsignificant difference) is the relative position of a cycle within a range. This may imply we do not have sufficient data to uncover a complete pattern in the lower cycles. Something on the order of 70% of the responses fell in the highest log cycle of a range.

The results are not the same for comparisons involving multiple-digit ranges. There are 15 separate log cycles (7B through 15C) for the pure multiple-digit case. Hence, there are $(15 \times 14)/2 = 105$ possible comparisons among pairs of ranges, as indicated in the lower right part of Table 1. There are 11 nonsignificant comparisons out of a possible 105 or about 10%. These 11 cases do not reflect the finding that matching occurs between the highest cycles in the range. Rather, all 11 cases are where members of the pair have identical cycles (A, B, or C), but these still only represent about 15% of all the pairs possessing the same cycle label.

No such pattern occurs for the mixed cases which comprise $10 \times 15 = 150$ pairs (upper right section of Table 1), of which 24 or about 16% are nonsignificant. We see no simple pattern in these results, although they may represent a transition between the unit-digit and multiple-digit results. More complete data are needed before such speculation can be adequately tested.

Response Uncertainty. Yet another way to consider the results is suggested by information theory as applied previously (Baird, 1970a, 1970b, 1970c). In the two-stage version of this psychophysical theory, it is assumed there are two types of information relevant to judgments, perceptual and cognitive. Perceptual

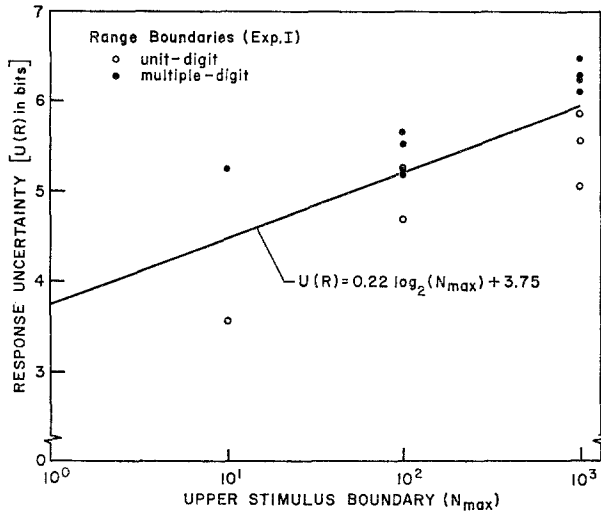


Fig. 3. Response uncertainty as a function of the log of the upper stimulus boundary used in Exp. I. Points are based on an evaluation of Eq. 1 for the group data for unit-digit (open circles) and multiple-digit (closed circles) ranges. The equation on the graph was determined by a visual fit of a straight line

information is determined by the attribute being judged and by the memory constraints imposed by the psychophysical method. Cognitive information is determined by response or output variables and is assumed to depend upon the ability to recall and use response categories (Baird, 1970c). Hence, it might prove fruitful to treat the cognitive factor in this theory as equivalent to response uncertainty. Such a distribution-free measure can be obtained by assuming S has a fixed number of response categories, each with their individual probabilities and that the group data from the present experiment validly represent these individual conceptions of the number scale. Then, because response uncertainty can be computed for each range, it is possible to determine whether the cognitive factor is constant or variable for different range sizes and locations.

The analysis was based on group data. First, the frequency of each response category used by S s was tabulated and divided by the total number of responses in order to obtain estimates of individual response probabilities. Second, Eq. (1) was applied to determine the average response uncertainty for each range:

$$U(R) = - \sum_{i=1}^M p(x_i) \log_2 p(x_i), \quad (1)$$

where $p(x_i)$ is the probability of the i th category, and M is the number of categories involved.

The results are presented in Fig. 3, where $U(R)$ in bits is plotted as a function of the upper stimulus boundary in the range (on a log scale). The open circles represent unit-digit ranges (1 through 6 in Table 1); the filled circles, multiple-digit ranges (7 through 15). In the latter, the data points have been shifted slightly to correspond with the x -ordinate locations of the nearest unit-digit ranges.

Once again (as in Table 1), we found that the lower stimulus boundary was considerably less important than the upper boundary. Moreover, response uncertainty is not constant for all ranges. The values for the multiple-digit ranges are higher than for the unit-digit ranges, and in both instances, $U(R)$ increases with increases in the magnitude of the upper boundary. The separate functions for unit and multiple digits may converge at some higher upper boundary as the lower boundary becomes less important in both cases. Fig. 3 shows that the number and/or probability of response categories is not fixed but varies systematically with stimulus and response range. Consequently, if $U(R)$ is taken as a measure of the cognitive factor, it is not a constant. One can write an equation to describe the relationship between $U(R)$ and the upper stimulus boundary (N_{\max}). A straight line was fit by eye to the data in Fig. 3 and the parameter values of the function determined.

$$U(R) = .22 \log_2 (N_{\max}) + 3.75 \quad (2)$$

Eq. (2) can be used to obtain an approximate value of the cognitive factor if one assumes that N_{\max} is the maximum response given with the method of magnitude estimation. In conjunction with previous theory (Baird, 1970 b), this indicates that the exponent of a power function (obtained by magnitude estimation) would decrease with increasing response range since the two-stage version of this theory states that $n = p/c$, where n is the exponent, p is the perceptual information associated with the judged attribute, and c is the cognitive information associated with the response attribute. With a constant p , the exponent n would be inversely related to c , which increases with response range according to Eq. (2). Such a prediction is in accord with empirical results (Baird, 1970 d; Teghtsoonian, 1971).

Experiment II

One major difference between the procedure of Exp. I and the method of magnitude estimation concerns the use of standards. In the stimulus range of Exp. I there were *two* boundaries delimiting the realm of possible responses. In magnitude estimation, *one* standard is usually employed, and *Ss* have the option of producing smaller or larger responses. If the results on number generation are to be generalized to situations involving magnitude estimation of more usual physical stimuli, it is important to show that the response biases obtained in Exp. I with two boundaries occur as well when there is only one boundary. Therefore, Exp. II was designed to gather more open-ended numerical responses with a single standard located at the lower boundary. In addition, we were not as concerned here with the particular responses as with the spacing between them for different boundaries. We wanted further evidence on the possibility that the spacing between responses is a constant proportion of the standard. Because of this orientation, *Ss* were allowed to generate *any* numbers (larger than the standard) they wished and were not told to predict responses of previous *Ss*. We fully expected this procedure to invite bizarre responses, and as it turned out, such expectations were confirmed. However, a sufficient amount of usable data was obtained to allow the conclusion that the spacing of responses for a single-boundary, open-ended task is essentially the same as for two-boundary ranges. In addition, because we forced *Ss* to order their responses according to relative magnitude, functions could be obtained between response order and magnitude (after Banks and Hill, 1974).

Method

Subjects. The participants were 73 male undergraduates enrolled in an introductory psychology course at Dartmouth College. None had served as *Ss* in Exp. I.

Stimulus Materials. As in Exp. I, the appropriate lower stimulus boundary was typed at the top of individual sheets of paper which were stapled together in a booklet for each *S*. No

upper boundary was given. The order of the sheets was randomized separately in each case. Sixteen spaces were marked on a sheet with a lower boundary number occupying the top space. Fifteen responses were then required for each of seven lower boundaries: 1, 5, 10, 12, 100, 211, and 1000. At this point in the experimental series, we were content to study boundaries containing one significant digit, and consequently, less emphasis was placed on having a full complement of multiple-digit boundaries.

Procedure. The experiment took place in the auditorium used in Exp. I. All *Ss* participated at the same session. They were told to write down 15 numbers greater than the boundary number at the top of each sheet. The only other constraint on the 15 numbers was that they be arranged down the sheet in a sequential order of magnitude.

Results and Discussion

Here we are interested in analyzing the results from a somewhat broader viewpoint than the one adopted in Exp. I. As in the latter, however, the relative error (standard deviation divided by the mean) is a good index of a pattern discernible in the findings. Because *Ss* were allowed to use numbers of any magnitude (no upper boundary), the arithmetic mean and standard deviation are inappropriate for describing the group data but appropriate for individual *Ss* since each produced 15 numbers for each boundary. In other words, the absence of an upper response boundary permitted more opportunity for individual differences to be manifested here, as contrasted with the situation in Exp. I. Consequently, the analysis was carried out at the individual level.

Fig. 4 shows relative error as a function of the mean for the unit-digit boundaries. Each data point in Fig. 4 is from one *S* on one boundary. Some of the responses were extremely large, so several restrictions were placed on the data finally treated. (1) All responses for a single boundary were excluded if at least one overtaxed the calculating ability of Dartmouth's Honeywell G635 computer. This meant that not all ranges were included for all *Ss*. (2) Once criterion 1 was passed, mean responses larger than 10^5 were also excluded. There were only 15 points between 10^5 and 10^9 , and they continue the trend (although flatter) laid down in Fig. 4.

The general trend in Fig. 4 is similar to that found in Exp. I (Fig. 1). Since the overall level of the data points is essentially parallel to the x -axis, the results support the constant-ratio model and contradict the constant-interval model. The four open circles to which the arrows point represent special cases, namely, instances where *S* responded with multiples of the boundary, that is, 1, 2, 3, . . . 16; 10, 20, 30, . . . 160, etc., thereby supporting perfectly the constant-ratio model. For example, the mean of the numbers 2 through 16 is 9.0 with a relative error (standard deviation divided by the mean) of 0.48. There are 62 measurements of this type. However, within a single boundary, the relative error rises faster than is suggested by the overall trend. The dispersion of these points for a particular boundary is an indication of individual differences. The comparable data for the 5 boundary is very similar to the results in Fig. 4, whereas there is considerably more dispersion in the data for the 12 and 211 boundaries.

There are substantial individual differences within a boundary condition, but this does not mean that similar variability exists for an *S*'s responses for different boundaries. The constant-ratio model would be bolstered by evidence indicating that the relative positions of data points for different *Ss* in the group were the same for each boundary. To test this hypothesis, correlations were determined

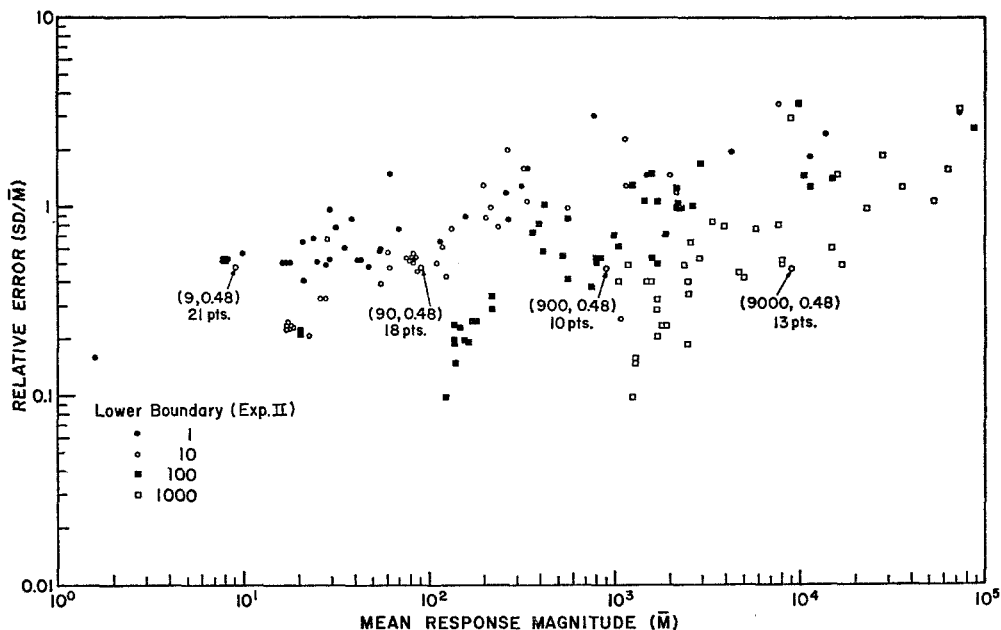


Fig. 4. Relative error as a function of mean response magnitude. The coordinates are logarithmic. Points are based on the data from individual *Ss* in Exp. II. The arrows point to cases in complete support of the constant-ratio model. The numbers there refer to the number of means obtained at that position. Other points on the graph for which more than one mean was obtained are not indicated. For more details, see the text

Table 2. Correlation coefficients (*r*) from Exp. II. Results are based on the relative errors of individual *Ss*

		Lower Boundary						
		1	5	10	12	100	211	1000
Lower Boundary	1	1.0	.03	.25*	.26*	.32*	.20	.08
	5		1.0	.21	.03	.28*	.35*	.48*
	10			1.0	.37*	.42*	.36*	.26*
	12				1.0	.39*	.17	.25*
	100					1.0	.36*	.38*
	211						1.0	.16
	1000							1.0

* $p < .05$, $df = 54$, one-tailed test.

among the seven boundary conditions, where the data pairs were the two relative errors obtained from the same *S* for two boundaries. Not all *Ss* had usable data for all the boundaries, and 17 were excluded on this basis, leaving 56 *Ss* to be analyzed. The correlation matrix is given in Table 2. We note, first, that all correlations are positive, and 66% of the entries are significant ($p < .05$, $df = 54$, one-tailed test).

Table 2 presents strong evidence that *Ss* gave responses in a characteristic fashion for each boundary. Such results have implications for recent discussions of individual differences in ratio estimation. In particular, they suggest that stable individual differences can be found in psychophysical judgments, in agreement with statements by Ekman *et al.* (1968) and in disagreement with the conclusions implied by Teghtsoonian and Teghtsoonian (1971), who did not find significant correlations among individual exponents obtained successively over time periods stretching from several days to one year⁴.

The Order Statistic. Another perspective on the data is secured by looking at the relation between rank order of response and response magnitude. The *Ss* listed their responses in a sequential rank order beginning with a number larger than the standard. The nature of this relation is helpful in deciding upon the merits of different psychophysical functions (e.g., log, power). More specifically, we assume that equal spacing between ranks represents equal spacing between perceptual scale values (Fechner's (1907) assumption) and that the psychophysical function is between these ranks and the numbers generated for each. This approach to number scaling was first used by Banks (1973).

In order to bring all the data within manageable and comparable limits, the numbers in each range for each *S* were first multiplied so that their geometric mean was equal to 100 times the standard. Subsequently, the geometric means were determined for the group at each position (rank value) within each range⁵. The results are presented in Fig. 5. There we have plotted the relation between rank order on a linear scale and geometric mean on a logarithmic scale. All of the functions are linear throughout most of their course, although some have a slight hook at the lower and upper ends, leading to the visual impression of an S-shape. This general characteristic was found also by Banks (1973) and by Banks and Hill (1974). If one eliminates the smallest and largest points from each function, the remainder form a clear linear relation⁶.

The slopes of the functions are not the same for all standards. The lowest slope occurs for the 1 standard, the highest for the 211 standard. One consistent finding here is that slopes for the unit-digit ranges (excluding the 5 standard) are lower than those of multiple-digit ranges (including the 5 standard). This implies that the intervals between response categories are smaller for the multiple-digit standards, a conclusion suggested also by the uncertainty analysis (Fig. 3).

We also attempted to fit power functions to the data. In a log-log plot, the relation between rank order and response magnitude was curvilinear throughout, even with the lowest and highest points eliminated. The shape of these functions was concave downward in a log-log plot as they should be if good linear functions obtain on semilog paper (Fig. 5).

In summary, the order statistic implies that Ekman's formulation (1964) is quite acceptable in that the relation between physical and perceptual numbers is

⁴ In an unpublished study on line length we obtained a significant correlation of .7 among exponents for 24 *Ss* retested after one year. This continuum was also used by the Teghtsoonians. However, they only had 10 *Ss* in a group, an inadequate number (in our opinion) for a correlational analysis of this type.

⁵ The results were essentially unchanged when we eliminated all responses greater than 10°.

⁶ We were able to obtain good linear fits by the method of least squares.

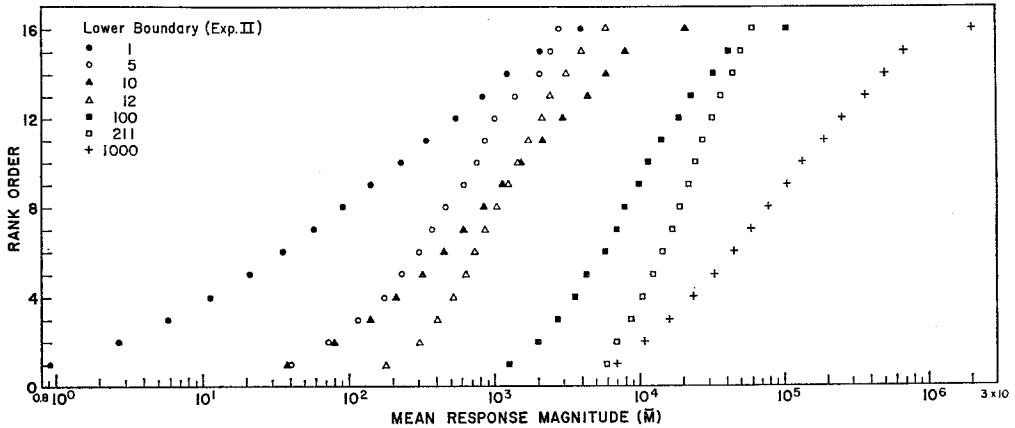


Fig. 5. Rank order of response as a function of the logarithm of the geometric mean based on group data from Exp. II. For more details, see the text

logarithmic throughout most of the range. Moreover, since the slopes in Fig. 5 depend on the standard, exponents would also depend on the type of standard or on the range of responses induced by a standard. If all the functions in Fig. 5 had the same slope, it would mean that response magnitude was perfectly related to the size of the standard, and hence, response bias would not be introduced by using a variety of standards. However, such a bias is an empirical fact (Engen and Ross, 1966). More specifically, multiple-digit standards and single-digit standards *other* than 1, 10, 100, etc., should produce lower exponents in magnitude estimation tasks since, in the exponent $n = p/c$, the value of the cognitive factor (c) would be larger for "odd" standards (Baird, 1970b). Another way to look at it is that one would not advance as far along the response scale for "odd" standards because there are more steps (scale values) per unit change in intensity along the matched attribute scale. This prediction is implied as well by the results of Wong (1963).

General Overview

We are now in a position to appreciate, along with Francis Galton (1880), the remarkably complex views people have of the mathematical number scale. Despite this complexity, several invariant patterns emerge from this study. It is convenient to separate these patterns into those found at a molar, molecular, and supermolecular level of analysis. Invariances discovered at one level need not be found at another, although this fact does not diminish the importance of either in our effort to understand the processes by which people generate numbers.

Molar Pattern. The most general overview can be obtained by looking at the results on relative error (Exps. I and II) and the order statistic (Exp. II). The relative error is reasonably constant across changes in the mean response magnitude for both unit- and multiple-digit ranges in Exp. I (Fig. 1) and for unit-digit lower boundaries in Exp. II (Fig. 4). This suggests that the spacing between response categories measured in the corresponding physical units is a positive,

linear function of the mean of the responses. If this were indeed true throughout the number scale, we would expect to find a logarithmic relation between response order (rank) and magnitude for the conditions of Exp. II in accord with Ekman's (1964) formulation. The results in Fig. 5 confirm this expectation for much of the relevant range (excluding the smallest and largest means in the series).

Molecular Pattern. Although the order statistic suggests that a logarithmic function holds over most of the range, a more detailed analysis reveals that this effect is due largely to differences between log cycles and that within a cycle, a fairly constant interval obtains between major response categories. Fig. 2 (top) highlights the dominance of single-significant-digit responses. The latter finding implies that the spacing between categories increases linearly with increases in the magnitude of the cycle. Considering Fig. 3, we conclude that a constant ratio is generally maintained between responses from two log cycles but that within a cycle, interresponse interval is approximately constant.

One further effect apparent at this level analysis is the reliability of individual differences. The significant correlation coefficients in Table 2 (Exp. II) indicate that *S*s are consistent in their response patterns across stimulus ranges.

Supermolecular Pattern. A more careful examination uncovers two further facts which were not emphasized. (1) There are more significant digits used (resulting in more categories) for the higher log cycles. (2) There are more significant digits used for the multiple-digit ranges. This effect, though not pronounced, is implied by the rising relative error in Figs. 1 and 4 and by the uncertainty analysis (Fig. 3). Although not evident from Fig. 3, the response uncertainty [$U(R)$] is greater in higher log cycles. The lower boundary is relatively unimportant since approximately 70% of the responses fall within the highest log cycle induced by the stimulus conditions.

The next paper in this series focuses on theoretical models developed to clarify the judgment processes underlying the major results of this study.

References

- Attneave, F.: Perception and related areas. In: Koch, S. (Ed.): *Psychology: A study of a science*, Vol. 4, pp. 619–659. New York: McGraw-Hill 1962
- Baird, J. C.: A cognitive theory of psychophysics. I. Information transmission, partitioning, and Weber's law. *Scand. J. Psychol.* **11**, 35–46 (1970 a)
- Baird, J. C.: A cognitive theory of psychophysics. II. Fechner's law and Stevens' law. *Scand. J. Psychol.* **11**, 89–102 (1970 b)
- Baird, J. C.: Information processing in alternative visual spaces. Chapter in: Baird, J. C. (Ed.): *Human space perception: Proceedings of the Dartmouth conference*. Psychonomic Monogr. Suppl. **3**, (Whole No. 13) (1970 c)
- Baird, J. C.: *Psychophysical analysis of visual space*. London: Pergamon 1970 d
- Baird, J. C., Lewis, C., Romer, D.: Relative frequencies of numerical responses in ratio estimation. *Percept. Psychophys.* **8**, 358–362 (1970)
- Baird, J. C., Kreindler, M., Jones, K.: Generation of multiple ratio scales with a fixed stimulus attribute. *Percept. Psychophys.* **9**, 399–403 (1971)
- Banks, W. P.: A new psychophysical ratio scaling technique: Random production. *Bull. Psychonom. Soc.* **1**, 273–275 (1973)
- Banks, W. P., Hill, D. K.: The apparent magnitude of number scaled by random production. *J. exp. Psychol. Monogr.* **102** (No. 2) (1974)
- Curtis, D. W.: Magnitude estimations and category judgments of brightness and brightness intervals: A two-stage interpretation. *J. exp. Psychol.* **83**, 201–208 (1970)

- Curtis, D. W., Attneave, F., Harrington, T. L.: A test of a two-stage model of magnitude judgment. *Percept. Psychophys.* **3**, 25—31 (1968)
- Eagleston, O. W., Lipford, E. J.: A study of number choices. *J. genet. Psychol.* **31**, 129—133 (1944)
- Ekman, G.: Is the power law a special case of Fechner's law? *Perceptual and Motor Skills* **19**, 730 (1964)
- Ekman, G., Hosman, B.: Note on subjective scales of number. *Perceptual and Motor Skills* **21**, 101—102 (1965)
- Ekman, G., Hosman, B., Lindman, R., Ljungberg, L., Åkesson, C. A.: Interindividual differences in scaling performance. *Perceptual and Motor Skills* **26**, 815—823 (1968)
- Engen, T., Ross, B. M.: Effect of reference number on magnitude estimation. *Percept. Psychophys.* **1**, 74—76 (1966)
- Fechner, G. E.: *Elemente der Psychophysik*, Bd. II. Leipzig: Breitkopf und Hartel 1907
- Galton, F.: Visualized numerals. *J. anthropol. Inst.* **10**, 85—97 (1880)
- Heywood, S.: The popular number seven or number preference. *Perceptual and Motor Skills* **34**, 357—358 (1972)
- McGill, W.: The slope of the loudness function: A puzzle. In: Gulliksen, H., Messick, S. (Eds.): *Psychological scaling: Theory and applications*. New York: Wiley 1960
- Rosner, B. S.: The power law and subjective scales of number. *Perceptual and Motor Skills* **21**, 42 (1965)
- Ross, B. M., Engen, T.: Effects of round number preferences in a guessing task. *J. exp. Psychol.* **58**, 462—468 (1959)
- Ross, S., Kohl, D. M.: Perceptual factors in number choices. *J. genet. Psychol.* **39**, 39—47 (1948)
- Rule, S. J.: Equal discriminability scale of number. *J. exp. Psychol.* **79**, 35—38 (1969)
- Stevens, S. S.: On the operation known as judgment. *Amer. Scientist* **54**, 385—401 (1966)
- Teghtsoonian, M., Teghtsoonian, R.: How repeatable are Stevens' power law exponents for individual subjects? *Percept. Psychophys.* **10**, 147—149 (1971)
- Teghtsoonian, R.: On the exponents in Stevens' law and the constant in Ekman's law. *Psychol. Rev.* **78**, 71—80 (1971)
- Wagenaar, W. A.: Generation of random sequences by human subjects.: A critical survey of the literature. *Psychol. Bull.* **77**, 65—72 (1972)
- Winick, C.: Preference for individual digits. *J. genet. Psychol.* **67**, 271—281 (1962)
- Wong, R.: Effect of the modulus on estimates of magnitude of linear extent. *Amer. J. Psychol.* **76**, 511—512 (1963)
- Yule, G. U.: On reading a scale. *J. roy. statist. Soc.* **90**, 570 (1927)

Dr. John C. Baird
 Department of Psychology
 Dartmouth College
 Hanover, New Hampshire 03755
 USA