Psychol. Res. 38, 81 - 95 (1975)

Psychophysical Study of Numbers

II. Theoretical Models of Number Generation*

ELLIOT NOMA** and JOHN C. BAIRD

Department of Psychology, Dartmouth College, Hanover, New Hampshire 03755, USA

Received April 30, 1974/ Revised June 20, 1975

SUMMARY. This paper develops three theoretical models to predict the numbers generated by Ss in an experiment described by Baird and Noma, 1975, Exp. II. The models (digit, base, and quarter) are each grounded on different assumptions about the process underlying number generation without the constraints of physical stimuli usually present in psychophysical tasks. Each of the models proved applicable to a restricted subrange of the physical continuum from 1-1000. A combination of models seems necessary to adequately predict number generation.

INTRODUCTION

Francis Galton (1880) clearly understood that a person's conception of the mathematical number scale provides fascinating but complex material for theoretical study. In the initial part of this work (Baird and Noma, 1975), it was shown that Galton's view of the problem was somewhat more realistic than the views held by some modern theorists interested in perception of the number continuum. Most importantly, we found that the perception of numbers is not a simple function of physical scale values, although Ekman's (1964) formulation of a logarithmic relation

---

* This article is the second in a five-part series. Reprint requests should be sent to John C. Baird.

** Presently at the University of Michigan.

between perceived and physical number is reasonably correct at a molar level of analysis. However, the details of the function are considerably more intricate. In tasks requiring Ss to generate numbers within boundary values (e.g., 1-100, 10-1000), the probability of number occurrence depends upon several factors, including the size of the boundaries, the location of the number within the range, and whether the number is a multiple of certain special integers, such as 1, 5, and 10 (Baird and Noma, 1975).

The purpose of this article is to develop several quantitative models to predict the numbers (and their frequencies) produced under the conditions tested in these experiments. Succeeding parts of this study deal with methodological applications of numbers as stimuli in standard psychophysical tasks (Weissmann, Hollingsworth, and Baird, 1975), as well as with the broader implications of this research for theories of scaling involving other stimulus attributes (Baird, 1975, a,b).

There are two aspects to this problem. The first concerns predicting the probability that a number falls within different $\log_{10}$ cycles (e.g., 1-10, 10-100, 100-1000). That is, when Ss are asked to generate numbers within boundaries enclosing two or more log cycles, the probability of number production among cycles is not equal. This aspect is considered to be peripheral to the major problem of predicting frequencies within a log cycle although we attempt to include both aspects in the models.

Cycle Selection. Once a stimulus range is given, there is a certain probability that a number will be chosen within each of the appropriate log cycles. A model of cycle selection can specify these probabilities, although at present, we have no adequate explanation for them.

We start with a few definitions. A number (N) is defined as falling within a cycle if $10^n \le N < 10^{n+1}$. Since the highest cycles in the ranges are statistically indistinguishable (Baird and Noma, 1975; Komogorov-Smirnov tests, Tab. 1), we define the upper complete cycle of a stimulus range as R_n. The ranges under consideration will be restricted to "unit digit" cases: 1-10, 1-100, 1-1000, 10-100, 10-1000, and 100-1000. These ranges cover at least one complete cycle and one number from the next highest. For example, for the range 10-100, a response in the upper cycle is $10^1 \le N < 10^2$, but responses of 10^2 were allowed, so these are considered to fall within the next highest cycle $10^2 \le N < 10^3$. This applies to the stimulus range (R). A similar definition is assumed for response cycles (C).

Turning now to some data, the relative frequency of a number falling within different response cycles (C) can be calculated for each of the stimulus ranges. Cycle C_n is always the highest cycle; therefore C_{n+1} covers the relative frequency of a response equal to the upper boundary (10, 100, or 1000) in the unit digit cases. We can simplify matters considerably by redistributing these frequencies among the remaining response cycles. That is, assuming responses to the upper boundary were disallowed, how would the remaining frequencies redistribute? This can be found (with reasonable assumptions) for each of the cycle frequencies $p(C_x)$ by applying Eq. 1

$$\frac{p(C_x)}{p(C_n) + p(C_{n-1}) + p(_{n-2})}, \tag{1}$$

where $p(C_x)$ is successively set equal to $p(C_n)$, $p(C_{n-1})$, and $p(C_{n-2})$. The results from Eq. 1 are given in Tab. 1, examination of which suggests that $p(C_n)$ is quite similar for each of the ranges containing more than one cycle. The average value of $p(C_n)$ over the three multiple-cycle ranges is .7. For two-cycle ranges, this means that $p(C_{n-1}) = .3$. For the three-cycle range 1-1000, we assume from the data that $p(C_{n-1}) = .1$ and $p(C_{n-2}) = .2$ (see Tab. 1) are reasonable values.[1]

This allows us, then, to formulate a descriptive model of response cycle selection conditional upon stimulus range.

1-10, 10-100, 100-1000	$p(C_n) \mid R_n = 1.0$
all multiple-cycle ranges	$p(C_n) \mid (R_n+R_{n-1} \ldots R_{n-x}) = \alpha = .7$
1-100, 10-1000	$p(C_{n-1}) \mid (R_n+R_{n-1}) = 1-p(C_n)$
1-1000	$p(C_{n-1}) \mid (R_n+R_{n-1}+R_{n-2}) = \gamma = .1$
	$p(C_{n-2}) \mid (R_n+R_{n-1}+R_{n-2}) = 1-\alpha-\gamma$

Table 1. Relative frequency of response values. Data given for different log cycles for each of six stimulus ranges used by Baird and Noma, 1975

		Log Cycles			
		C_{n+1}	C_n	C_{n-1}	C_{n-2}
Range	1 -10	-	1.0	-	-
	10 -100	-	1.0	-	-
	100-1000	-	1.0	-	-
	1 -100	-	0.79	0.21	-
	10 -1000	-	0.64	0.36	-
	1 -1000	-	0.68	0.12	0.20

[1] $p(C_x)$ will henceforth be referred to as a probability.

These descriptive equations are clearly limited by a lack of theoretical understanding about the psychological variables important in selection of a cycle, but they can serve as an initial step toward further work. Larger stimulus ranges must be tried before further generalizations are attempted.

Number Selection. Three models of number selection are suggested here: the *digit* model, the *base* model, and the *quarter* model. We will first describe each and then show the degree of correspondence between their predictions and the empirical data obtained for the six unit-digit ranges. None of the models is completely satisfactory for all ranges, although a combination of the base and quarter model shows the most potential for quiding future work.

The Digit Model

This model gives the probability of selecting one or more significant digits within each of the stimulus ranges. The actual selection of a nonzero digit (1, 2,. . . 9) is assumed to be a random process.

Tab. 2 gives the relative frequencies for one, two, and three significant digits occurring in a number generated by Ss within each of the stimulus ranges (R) once the response cycle is specified (C). These frequencies are adjusted values based on Eq. 1 in order to eliminate responses equal to the upper stimulus

Table 2. Relative frequency of significant digits in response numbers. Data based on six stimulus ranges used by Baird and Noma, 1975

			Stimulus Range (R)					
		Significant Digits	1-10	10-100	100-1000	1-100	10-1000	1-1000
Response Range (C)	C_n	1	0.97	0.68	0.64	0.59	0.64	0.65
		2	0.03	0.32	0.22	0.41	0.17	0.15
		3	-	-	0.14	-	0.19	0.20
	C_{n-1}	1				0.96	0.58	0.54
		2				0.04	0.42	0.45
		3				-	-	-
	C_{n-2}	1						0.99
		2						0.01
		3						-

boundary. From these data, it can be seen that the probability of obtaining one significant digit is predominant and similar for all response cycles, except for cases involving the range $10^0 \leq N < 10^1$. In the latter instance, the probability is close to 1.0 that numbers will contain one significant digit. Excluding these cases, the mean relative frequency is .61. Formalizing matters, we conclude that the probability of one significant digit (D_1) can be found by Eqs. 2 and 3.

$$p(D_1) \mid [C_n = (10^0 \leq N < 10^1)] = 1.0 \tag{2}$$

$$p(D_1) \mid (C_n;\ C_{n-1};\ \ldots\ C_{n-x}) = \beta = .6;\ C_n \neq (10^0 \leq N < 10^1) \tag{3}$$

Then, the probability of two significant digits (D_2) for response cyles C_n and C_{n-1} (associated with stimulus ranges 1-100 and 10-1000) is seen to be

$$p(D_2) \mid (C_n;\ C_{n-1}) = 1 - \beta;\quad C_{n-1} \neq (10^0 \leq N < 10^1) \tag{4}$$

The determination of D_2 for the three response cycles ($C_n;C_{n-1};C_{n-2}$) associated with the ranges 100-1000, 10-1000, and 1-1000 requires one further assumption. Namely, we assume that the probability of adding a significant digit is always equal to ($1-\beta$) times the total available numbers at that point. This assumption does little violence to the empirical data and reduces the free parameters in the model. Therefore,

$$p(D_3) \mid (C_n;\ C_{n-1};\ C_{n-2}) = (1 - \beta)^2;\ C_{n-2} \neq (10^0 \leq N \leq 10^1) \tag{5}$$

Finally, since the probabilities for one, two, and three significant digits must add to 1,0,

$$p(D_2) \mid (C_n;C_{n-1};C_{n-2}) = 1-p(D_1)-p(D_3) = 1-(\beta) - (1-\beta)^2 = \beta - \beta^2. \tag{6}$$

Eqs. 2 through 6 constitute a model for selection of significant digits within the constraints provided by the stimulus and response ranges given in Tab. 2. The model has one free parameter (β). The selection of a specific digit (1 . . . 9) is assumd to be a random process.

The Base Model

The empirical data suggest rather strongly that Ss generate certain numbers much more frequently than others and that these "preferred" numbers tend to be multiples of 1, 10, 100, 5, and 50. The base model elaborates on this theme by claiming that these are the only numbers worth considering when describing the response distribution.

Dealing first with multiples of 1, 10, and 100, the results can be described by a process functioning according to a mathematical base 10 system, where only one significant digit is used. We will refer to these responses as "preferred numbers" (N). The values for N can be found by applying a single exponential function:

$$N = kb^n , \tag{7}$$

where b is the base 10, n is the place integer (0, 1, 2, . . . etc.), and k is the category integer ranging from 1 to b - 1.

Preferred numbers generated by a base 10 system are clearly the most prevalent in the data (Baird and Noma, 1975, Fig. 2). Multiples of 5, 50, and 500 are also important, although not at equal strength for all multiples. In fact, a closer look at the individual multiples suggests that numbers such as 15, 25, 75, and 250 are much stronger than multiples such as 35, 85, 140, or 260. These preferred multiples suggest that a base 5 system is operating here in addition to base 10. If b = 5 in Eq. 7, the preferred numbers from this system can be obtained. Examples of preferred numbers for bases 10 and 5 are given in Table 3. The important aspect of the base system for our purposes is its generation of selected multiples. The actual base notation is unimportant. Therefore, entries in the table represent an evaluation of Eq. 7 in decimal notation for both bases. Assuming a preferred number can be obtained from either of these two bases and that outputs from two bases do not add, we can write out the transformation for any region of the number continuum. For instance, selecting the range of numbers from 1 to 1000, we have the following preferred states written in base 10 notation: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15*, 20, 25*, 30, 40, 50, 60, 70, 75*, 80, 90, 100, 125*, 200, 250*, 300, 375*, 400, 500, 600, 625*, 700, 800, 900, 1000. The asterisk indicates integers generated by base 5 alone. It is the contention of the base model that these numbers represent the major share of generated responses. Moreover, we assume that each preferred number is equally likely.

Table 3. Preferred states generated by base 10 and base 5. Entries are in decimal notation.

	b_n					b_n			
	10^3	10^2	10^1	10^0		5^3	5^2	5^1	5^0
1	1000	100	10	1	1	125	25	5	1
2	2000	200	20	2	2	250	50	10	2
k ⋮	⋮	⋮	⋮	⋮	k 3	375	75	15	3
9	9000	900	90	9	4	500	100	20	4

The Quarter Model

Although the base model captures the preferred numbers generated from 1 to 100, it does not seem as applicable to numbers greater than 100, at least for base 5. For example, this base system predicts that numbers such as 375 and 625 are important, and they clearly are not, either in our experiments or in psychophysical studies (Baird, Lewis and Romer, 1970).

The quarter model maintains the importance of 1, 10, and 100 but not necessarily within the context of the base model. These numbers are simply preferred multiples. In addition to these multiples, the model assumes that beyond 10, a log cycle is divided into quarters to produce the further preferred numbers 25,50,75,100,250,500,750, and 1000. These quarter values and the multiples of 1,10 and 100 are then weighted differentially. Finally, the model assumes additivity. These weightings are then used to predict probabilities of occurrence for each of the numbers falling within a particular stimulus range, The probability of a generated value falling within different log cycles is predicted by the same cycle selection model used for the digit and base models.

Kolmogorov-Smirnov Tests

The adequacy of the digit, base, and quarter models was determined by Kolmogorov-Smirnov tests of the differences between the theoretical and empirical distributions for each of the six unit-digit ranges reported in Baird and Noma (1975). This test considers the maximum absolute difference between the theoretical ($F_M(x)$) and empirical ($F_S(x)$) relative frequency distributions:

$$D = \max |F_M(x) - F_S(x)| \tag{8}$$

With the value of D and a good approximation for a continuous distribution from the large number of responses, the Kolmogorov-Smirnov test of goodness of fit may by used.

The parameter values for log cycle selection were identical for the digit and quarter models and were taken to be those determined from the data summary given in Table 1. That is, $\alpha = .7$, and $\gamma = .1$. Slight iteration of α provided better fits for the base model, and for these tests, $\alpha = .78$, and $\gamma = .06$.

For the digit model $\beta = .6$ for selection of significant digits, and for the quarter model the multiples of 1, 10, and 100 were weighted by 1.0, while the quarters were weighted by .5. These weights were determined to be optimal (by inspection) across stimulus ranges, as determined by iteration procedures.

The results of the Kolmogorov-Smirnov tests are presented in Fig. 1 through 5 for five ranges, with the exception of 1 to 10, for which all models provided an excellent fit. In each figure, it is possible to assess the degree of fit between the empirical data (presented in the lower part of the figure) and the theoretical models throughout the stimulus range (omitting the upper boundary, 10,100, or 1000). The upper three sections of each figure show the differences in cumulative relative frequency distributions for each of the models. The solid horizontal lines through zero indicate perfect agreement between empirical and theoretical data; the upper and lower horizontal lines are the boundaries (plus and minus) of nonsignificant differences ($p < .05$). The obtained differences are shown by the irregular continuous curves. The way to read these graphs is as follows: The predictions of the model are significantly different from the empirical data if the curve lies outside the boundaries in either a positive or negative direction for any point along the x-axis. This type of display allows one to specify more exactly the regions where the models have difficulty, an advantage clearly lost when one reports only the maximum difference used to determine statistical significance. In particular, negative deviations from zero indicate that the cumulative frequency of the empirical distribution up to x was greater than that of the theoretical distribution (Eq. 8). The opposite is of course true for positive deviations.

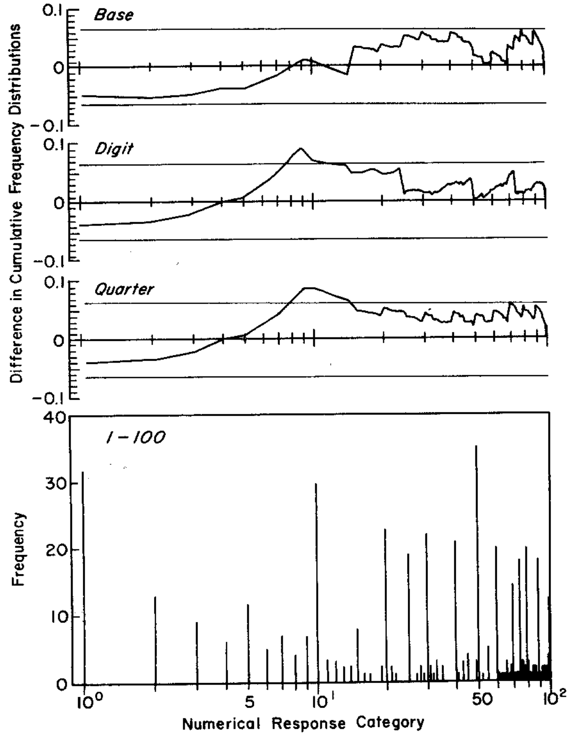

Fig.1. Relationship between empirical data and theoretical models for the stimulus range 1-100. Bottom part of the figure gives empirical results. Top sections present results of Kolmogorov-Smirnov tests (Eq. 8) for the base, digit, and quarter models. Nonsignificant differences between the two cumulative frequency distributions are indicated whenever the obtained curves remain within the positive and negative horizontal lines ($p = .05$). For more details, see the text.

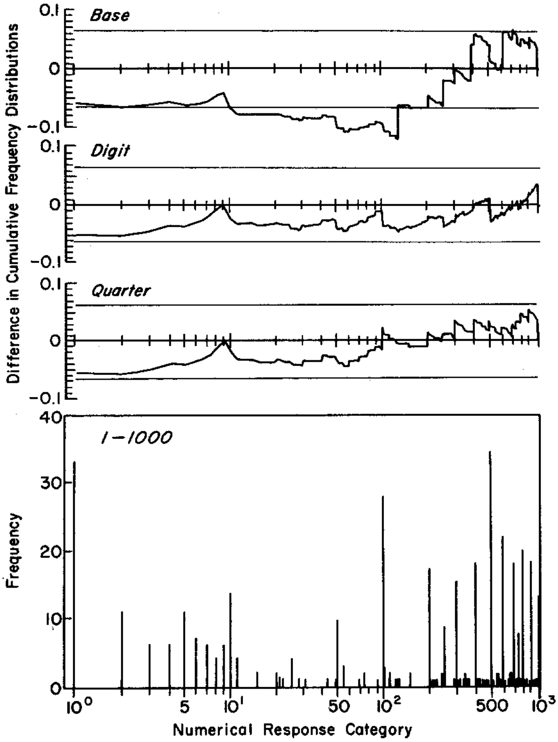

Fig. 2. Relationship between empirical data and theoretical models for the stimulus range 1-1000. Bottom part of the figure gives empirical results. Top sections present results of Kolmogorov-Smirnov tests (Eq. 8) for the base, digit, and quarter models. For more details, see the text and Fig. 1

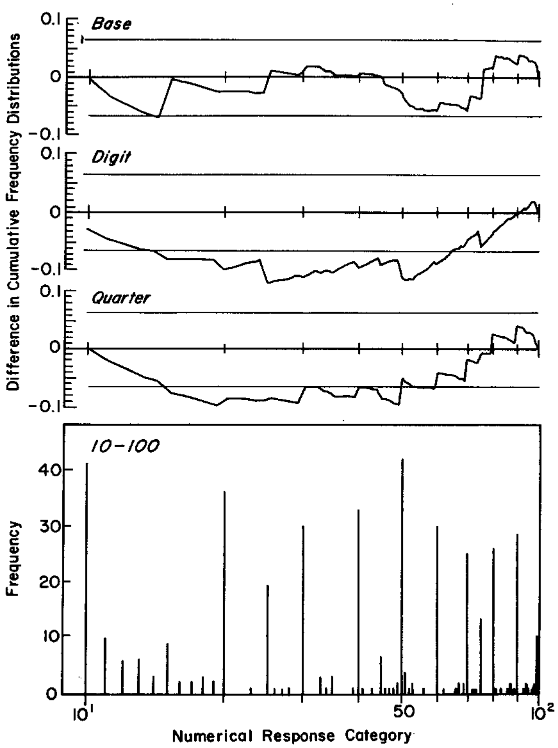

Fig. 3. Relationship between empirical data and theoretical models for the stimulus range 10-100. Bottom part of the figure gives empirical results. Top sections present results of Kolmogorov-Smirnov tests (Eq. 8) for the base, digit, and quarter models. For more details, see the text and Fig. 1

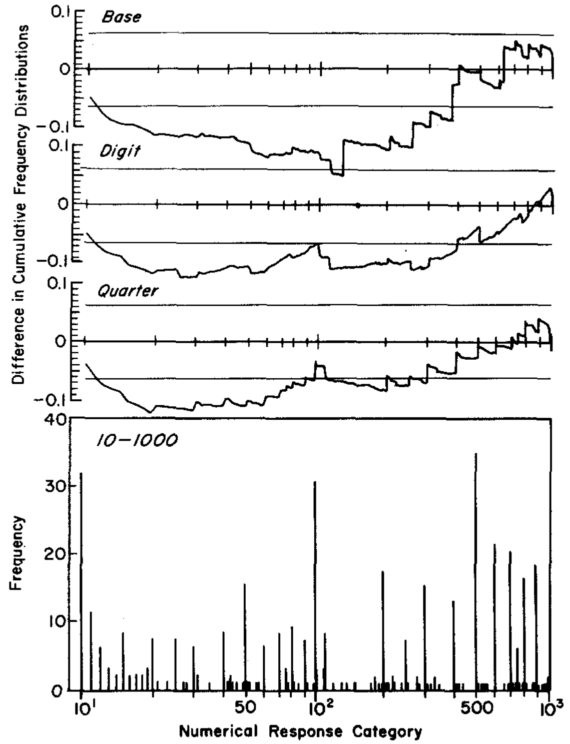

Fig. 4. Relationship between empirical data and theoretical models for the stimulus range 10-1000. Bottom part of the figure gives empirical results. Top sections present results of Kolmogorov-Smirnov tests (Eq. 8) for the base, digit, and quarter models. For more details, see the text and Fig. 1

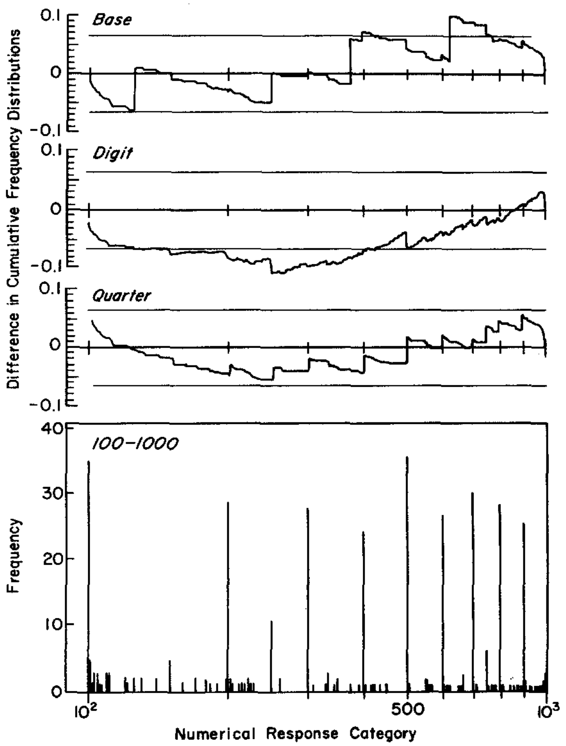

Fig. 5. Relationship between empirical data and theoretical models for the stimulus range 100-1000. Bottom part of the figure gives empirical results. Top sections present results of Kolmogorov-Smirnov tests (Eq. 8) for the base, digit, and quarter models. For more details, see the text and Fig. 1

Range 1-10. All models offer an excellent fit to the data for this range.

Range 1-100. Results are presented in Fig. 1. The predictions of the base model are not significantly different from the empirical results. Both the digit and quarter models exceed the positive boundary. Specifically, the theoretical distribution leading up to 10 becomes increasingly greater than the empirical function. However, the region falling outside the boundary is not extensive.

Range 1-1000. Results are presented in Fig. 2. Both the digit and quarter models are not significantly different from the empirical distribution, whereas the base model fails, primarily in the log cycle from 10 to 100, where empirical results are particularly scarce.

Range 10-100. Results are presented in Fig. 3. Both the digit and quarter models fail rather decisively, whereas the base model is adequate (one point is barely outside the boundary).

Range 10-1000. Results are presented in Fig. 4. All three models are inadequate. The empirical function has too many responses at the low end (e.g., 11, 12, 13, 14) and none of the models is able to recover from the initial negative drop induced by this situation.

Range 100-1000. Results are presented in Fig. 5. Only the quarter model handles this range satisfactorily. The digit and base models break down at different locations within the range.

In summary, each of the models seems applicable to different stimulus ranges. Although separate iteration of parameters for each range improves the fits, the overall pattern does not change dramatically. Furthermore, it appears that the base model is most applicable for numbers 1 to 100 (independent of the subrange selected), while the quarter model is more viable for numbers greater than 100. Hence, a combination of the major characteristics of both models would provide the best predictions for data generated in a variety of stimulus ranges. We will return to this possibility after presenting the three models' predictions of relative error (standard deviation divided by the mean) and uncertainty measures for each of the ranges. These predictions offer statistical summaries of the theoretical distributions used in the Kolmogorov-Smirnov tests (same parameter values).

Relative error. Fig. 6 shows the relation between theoretical relative error, as predicted by each of the models, and empirical relative error for each of the six ranges. In most cases, the theoretical values are less than the empirical ones, although agreement is fairly high for all three models. Considerably better fits can be obtained through individual iteration of

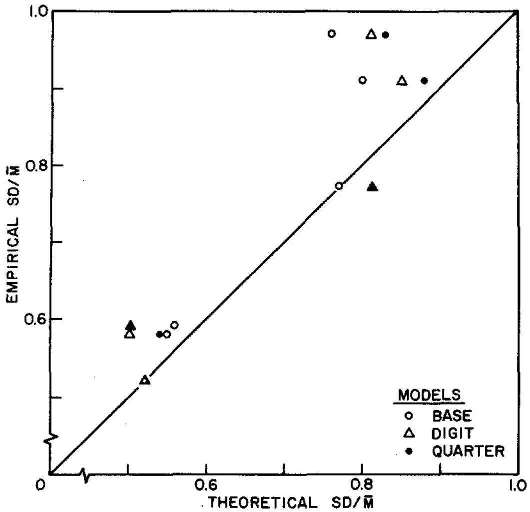

Fig. 6. Relationship between empirical standard deviation divided by the mean ($SD/\overline{M}$) and theoretical values obtained for the base, digit, and quarter models employing the parameter values stated in the text. Data are shown for each of six stimulus ranges.

parameters, but such optimization leads to less satisfactory agreement in terms of the Kolmogorov-Smirnov tests just discussed.

Uncertainty Measures. Response uncertainty measures were calculated according to Eq. 9 for each range on data generated by each model:

$$U = - \sum_{i=1}^{N} p(x_i) \log p(x_i), \tag{9}$$

where x_i was a single response category of the total N. These results are shown in Fig. 7, where empirical uncertainty is plotted against theoretical uncertainty. The base and quarter models have a limited number of response categories compared to the digit model, and this is reflected in the uncertainty measures. The uncertainties are higher for the digit model than for the empirical results, whereas the base and quarter models yield measures which are generally smaller than the empirical values. The existence of some low probability random categories (from the digit model?) to represent noise in the base and quarter models would bring theoretical predictions more in line with actual values.

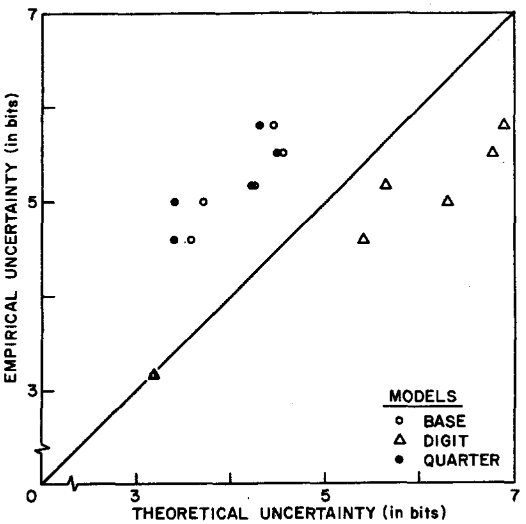

Fig. 7. Relationship between empirical uncertainty and theoretical values for the base, digit, and quarter models employing the parameter values stated in the text. Measures were obtained evaluating Eq. 9 for each of six stimulus ranges

OVERIEW[2]

The empirical and theoretical work presented in this and the first paper of the series permits us to devise a satisfactory picture of the number behavior of our subject population. The most obvious conclusion is that Ss prefer to use certain numbers at a much higher frequency than others, suggesting discrete steps in number preferences along the physical continuum. These preferred numbers are captured quite adequately by the three models described here, although each is most applicable to a different region of the continuum. Drawing upon the results presented here, as well as upon psychophysical studies requiring numerical responses (Baird et al., 1970), it is possible to provide a general description of number preferences (in terms of usage) for the range 1 to 1000. However, no single model seems able to handle all the results.

Assuming an equal weighting of log cycles in terms of the probability of number occurrence (this will no doubt depend upon the specific experimental conditions), we can describe number behavior for each log cycle separately. (1) For the range 1 to 9.9, the base model can be applied with only base 10 operating. Each preferred number is equally weighted in importance. (2) For the range 10 to 99, the base model is applicable using base 5 and base 10. As in the quarter model, these two bases can be

---

[2] All three models can also reproduce the function between rank order of response magnitude and the geometric mean described by Baird and Noma, 1975 (Fig. 5) and by Banks and Hill, 1974. Hence, this phase of the empirical results will not be discussed further.

weighted differently and their outputs assumed to be additive. Reasonable weights would appear to be 1.0 for base 10 and .5 for base 5. Hence, numbers such as 30, 40, and 60 are weighted by 1.0; numbers such as 25 and 75 are weighted by .5, and numbers such as 10 and 50 are weighted by 1.5. (3) For the range 100 to 999, the base 10 system continues to operate with a weighting factor of 1.0. However, the quarter model is used to obtain the numbers 250, 500, 750, which are weighted by .5. Numbers such as 100 and 500 receive strength from both the quarter and base 10 values and, hence, these receive the summated weight, 1.5. Another way to view the quarter values is that they represent 10 times the previous cycle of the base 5 system (i.e., $10(k5^2)$, where k = the integers 1 to 4). If the base 5 system were applicable for ranges extending past 100, the proper multiple would be 5 (i.e., $5(k5^2)$. Hence, it can be claimed that Ss are applying an inappropriate multiplier for numbers over 100.

Assuming equal weight for each log cycle and the foregoing model, we generated a frequency diagram for the continuum between 1 and 1000. This diagram is given in Fig. 8 and represents a prediction of the relation between the use of numbers and the physical scale (although random noise in the form of low frequency categories could be added). The close agreement between number generation data and data obtained from psychophysical studies (Baird et al., 197o) suggests that all such results are biased by relations similar to that given in Fig. 8 and would consequently have to be "corrected" in order to reveal the underlying scale appropriate for perception of the physical attribute (e.g., light, sound) under investigation.

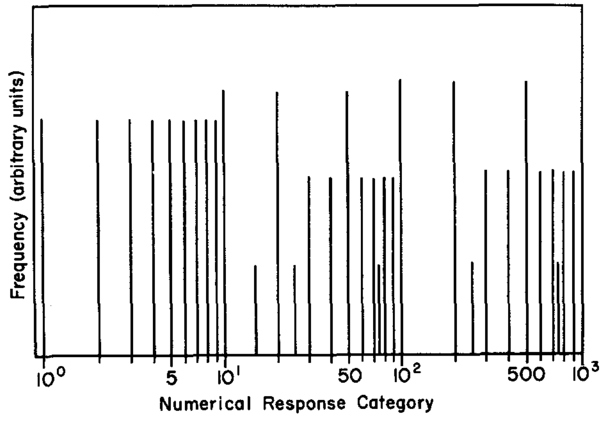

Fig. 8. Hypothetical frequency diagram for the range 1-999. Data were generated on the assumption of equal response frequencies among log cycles. The base model was employed for the range 1 to 100, whereas for numbers greater than 100, the quarter model was employed. Weights for particular types of response numbers are stated in the text

REFERENCES

Baird, J. C.: Psychophysical study of numbers: IV. Generalized preferred state theory. Psychol. Res. (1975a)

Baird, J. C.: Psychophysical study of numbers: V. Application of preferred state theory to matching functions. Psychol. Res. (1975b)

Baird, J. C., Noma, E.: Psychophysical study of numbers: I. Generation of numerical responses. Psychol. Res. 37, 281-297 (1975)

Baird, J. C., Lewis, C., Romer, D.: Relative frequencies of numerical responses in ratio estimation. Perception and Psychophysics 8, 358-362 (1975)

Banks, W. P., Hill, D. K.: The apparent magnitude of number scaled by random production. J. exp. Psychol. Monogr. 102, (No. 2) (1974)

Ekman, G.: Is the power law a special case of Fechner's law? Perceptual and Motor Skills 19, 73 (1964)

Galton, F.: Visualized numerals. Journal of the Anthropological Institute 10, 85-97 (1880)

Weissmann, S. M., Hollingsworth, S. R., Baird, J.C.: Psychophysical study of numbers: III. Methodological applications. Psychol. Res. (1975)