Vol. 18 Part 2 159—174
The British Journal of Mathematical and Statistical Psychology
November 1965

MAGNITUDE ESTIMATION OF HEAVINESS AND LOUDNESS BY INDIVIDUAL SUBJECTS: A TEST OF A PROBABILISTIC RESPONSE THEORY[1]

By R. DUNCAN LUCE and SUCHOON S. MO
University of Pennsylvania

One hundred magnitude estimates of heaviness were obtained for each of 18 weights from each of six *S*s and of loudness for each of 20 intensities from another six *S*s. The stimuli were spaced at approximately equal logarithmic intervals. The mean magnitude functions of individual *S*s appear, in general, to deviate systematically from power functions. The distributions of responses, normalized by the mean response, appear to be skewed, highly peaked, and to have high tails. For at least half of the *S*s, this response variability is no more than is needed to account for the errors in a simple two-stimulus two-response recognition experiment run under the same conditions. The data are examined in terms of a probabilistic response theory. Although the theory is clearly wrong in some of its details, certain of its qualitative features appear to be sustained.

1. INTRODUCTION

Several related methods, among them magnitude estimation (ME), magnitude production (MP), and cross-modality matching (CMM), have been proposed and used by S. S. Stevens (1957, 1960, 1961) to disclose the relation, or law, holding between the magnitude of sensation and a physical measure of the magnitude of the stimulus; such a relation is called *the psychophysical function.* These methods are viewed as ' direct ' in the sense that they completely bypass the usually untested, and to some extent untestable, assumptions of the indirect methods of Fechner, Thurstone, and others. The most striking and controversial outcome of these experiments is the form of the law to which they lend support: (approximate) power functions rather than the (approximate) logarithmic functions of the classical methods. Much attention has therefore been paid to the exact mathematical nature of the psychophysical function (Stevens 1957, 1960, 1961; Stevens and Galanter, 1957) and to its first and second order invariances (Rosner, 1961), but rather less effort has been devoted to a formal mathematical description of the several psychological processes underlying the observed responses.

To be more specific, little theory has been proposed to describe how the psychophysical function, biases and variability of individual *S*s interact to produce responses, and how the distorting effects of biases and variability can best be minimized. Stevens (1961) has advocated both averaging over *S*s and the use

[1] This work was supported in part by National Science Foundation Grant GB 1462 to the University of Pennsylvania. We are indebted to Drs. F. W. Irwin and S. S. Stevens for their criticisms.

of batteries of related experimental methods to eliminate the distortions and so obtain good estimates of the desired psychophysical function. His criteria for the adequacy of these procedures seems to be the smoothness, reproducibility and mathematical simplicity of the resulting psychophysical function. Several authors (Luce and Galanter, 1963; McGill, 1960; Pradhan and Hoffman, 1963) have expressed doubt whether these criteria alone are sufficient to conclude that the unwanted biases are largely eliminated rather than just smoothed. The usual approach to such a problem is to try to describe how the biases are generated in each *S*—i.e., to propose and test a theory that formulates the interaction of the psychophysical function with the various factors that may bias the responses—and then to determine mathematically schemes for processing the data that will markedly reduce, and perhaps minimize, the unwanted distortions. This necessarily requires detailed study of individual *S*s, not group averages. It is perfectly possible, as McGill (1960) has suggested, that the psychophysical function will turn out to be a ' personal equation ', in which case no averaging over *S*s is permissible.

At least two sources of potential bias must be considered. First, as many have repeatedly pointed out, ME admits of pure response biases which have something to do with the number habits of *S*s. For example, most *S*s seem to have not only whole number tendencies, but they round off to 5's and 10's. Such pure response biases are not restricted to ME, although they may be more extreme in that method; both MP and CMM probably involve biases that depend upon the details of the apparatus the *S* uses to generate his response. The major disadvantage of these methods may, in fact, be just this dependence on the unique configuration of specific pieces of laboratory equipment. This paper contributes little to the analysis of these biases, except to suggest that they may be fairly powerful. To find suitable combinations of experimental method and theoretical analysis sufficient to disentangle response biases from other determiners of the responses appears to be a formidable problem.

A second potential source of bias—the major focus of this paper—is the variability of the responses when the same stimulus is presented many times. To date, no studies have been concerned with the detailed nature of the response distribution of individual *S*s. This seems a bit puzzling, especially when we recall that the more classical methods of scaling have all rested in one way or another upon the existence of response variability. In our opinion, an adequate theory of ME must explicitly treat both the postulated psychophysical function, which presumably measures the growth of sensation, and the function representing response variability in *S*, which presumably measures the confusability of the stimuli for him. If variability depends only on confusability, we might hope to discover some relation between the variability observed in ME and that found in the responses of other methods, such as the absolute identification of one of a few stimuli.

One probabilistic response theory for ME (Luce and Galanter, 1963) begins precisely by taking the last observation seriously. That theory supposes, in effect, that ME is a continuous analogue of those simple recognition experiments in which one of a finite set of stimuli is presented on each trial and the *S* is required to identify which it is. Aside from the fact that the stimulus and response sets are (in principle) infinite in the one case and (small) finite sets in the other, the most crucial difference between the two experiments seems to reside in the relation existing between the stimuli and responses. In a recognition experiment a one-to-one correspondence between the stimulus and response sets, called the identification function (Bush, Galanter and Luce, 1963), is prescribed by *E*; it is known to both *E* and *S* and it establishes the meaning of each response to both. When the stimuli and responses are ordered, as is common in psychophysics, the identification function is usually monotonic increasing, although in principle it need not be. In ME a one-to-one correspondence between the stimulus and response sets, called a psychophysical function, is presumed to exist within *S*; it is unknown to *E* and, in a sense, to *S*. When the stimuli and responses are ordered, which again is usual, the psychophysical function is assumed to be monotonic increasing. Accepting the analogy as more than fortuitous, Luce and Galanter proposed as a possible theory for ME a continuous analogue of a so-called choice theory for recognition experiments (see Luce, 1963); this is outlined in some detail below.

The purpose of the two experiments reported here, the one concerned with heaviness and the other with loudness judgements, was to see whether or not (1) the estimated psychophysical functions for individual *S*s are power functions, as has been found for groups of *S*s, (2) the observed response distributions exhibit the peculiar peaked shape predicted by the theory, and (3) the variability of responses is of the same order of magnitude in ME as in a two-stimulus two-response recognition experiment run under the same conditions.

The theory proposed by Luce and Galanter is sufficiently general to be applicable to MP and CMM as well as to ME. It seems desirable, therefore, to collect sizable samples of data from each *S* using all three methods to see whether a single psychophysical function for each *S*, a single source of variability for each *S*, and a different response bias function for each method and each *S* are able to account for all of the data. This was not done in the present study because we were too uncertain about the theory and about the problems that might arise in collecting so much ME data from each *S* to justify such an extensive programme of research. Our data suggest, however, that it might be informative.

Theory

We assume that the stimulus and response sets are both ordered continua that are isomorphic to the positive real numbers ordered in the natural way. These numerical sets are labelled S and R, respectively. The unknown psychophysical function ψ, which is assumed to be a strictly monotonic increasing function from S into R, is interpreted as giving the magnitude of the sensory effect that results from a stimulus. Since the data strongly suggest that stimuli are confused with one another, it is assumed that the magnitude of this confusion can be measured by a function η in such a way that $\eta(s,s')$ is proportional to the probability density that, to S, stimulus s seems to be stimulus s'. We may transform this into a generalization function ζ' over responses by defining

$$\zeta'[\psi(s), \psi(s')] = \eta(s,s').$$

Within this model it is impossible to distinguish stimulus generalization from response generalization (this is generally so empirically, as was pointed out by Dawes, 1963). Finally, it is assumed that S may exhibit a bias to use certain responses more often than others; this tendency is represented by a function b from R into the real numbers such that $b(r)$ is proportional to the probability density of using response r independent of which stimulus is presented. The conditional response probability is then assumed to be given by

$$p(r \mid s) = \frac{\zeta'[\psi(s), r]b(r)}{\int_0^\infty \zeta'[\psi(s), x]b(x)dx}, \tag{1}$$

which is a continuous analogue of a class of models that have been used for discrete experiments (see Luce, 1963, p. 113–116).

The quantity estimated in ME is the expected response to a stimulus, i.e.,

$$E(r|s) = \int_0^\infty rp(r|s)dr. \tag{2}$$

It has been tacitly assumed by Stevens that $E(r|s)$ is proportional to the underlying psychophysical function $\psi(s)$, i.e., that there exists a constant μ such that

$$E(r|s) = \mu\psi(s). \tag{3}$$

It can be shown (Luce and Galanter, 1963, p. 287) that sufficient (but not necessary) conditions for eqn. (3) to hold are that the bias function be a power function $b(r) = \alpha r^{\gamma}$, where α and γ are constants, and that the generalization function ζ' depends only upon the ratio of its arguments, i.e., there is a function ζ of one variable such that

$$\zeta'(r,r') = \zeta(r'/r). \tag{4}$$

When $\gamma = 0$, i.e., when there is no response bias, as we will assume below, the constant of proportionality μ in eqn. (3) is the mean of ζ.

If, in addition to the assumption embodied in eqn. (4), we postulate that ζ is a continuous function and that it meets a condition of multiplicative independence, which has been found useful in experiments based on small finite sets of stimulus presentations which are selected from a one-dimensional stimulus continuum, namely, for all x and y such that either $0 \leqq x,y \leqq 1$ or $1 \leqq x,y < \infty$,

$$\zeta(xy) = \zeta(x)\zeta(y),$$

then it can be shown (*ibid.* p. 288) that ζ is of the following form:

$$\zeta(x) = \nu \begin{cases} x^{\delta}, & 0 \leqq x \leqq 1 \\ x^{-\epsilon}, & x \geqq 1, \end{cases} \tag{5}$$

where the normalizing factor is

$$\nu = \frac{(\delta+1)(\epsilon-1)}{\delta+\epsilon} \tag{6}$$

and $\delta > 0$, $\epsilon > 1$. If ζ is symmetric in the sense that $\zeta(x) = \zeta(1/x)$, then $\delta = \epsilon$. The data strongly reject this (inessential) assumption made by Luce and Galanter.

If $r(s)$ is a response to stimulus s, observe that from eqn. (3),

$$\frac{r(s)}{\psi(s)} = \frac{\mu r(s)}{E(r|s)}.$$

So if ζ^* is the distribution of $r(s)/E(r \mid s)$, which we can estimate from our data simply by forming the frequency distribution of a response to a stimulus divided by the mean response to that stimulus, and if ζ is the distribution of $r(s)/\psi(s)$, then it is easy to see that

$$\zeta^*(x) = \mu\zeta(\mu x), \tag{7}$$

where, from eqns. (5) and (6), the mean μ of ζ is given by

$$\mu = \left(\frac{\delta+1}{\delta+2}\right)\left(\frac{\epsilon-1}{\epsilon-2}\right). \tag{8}$$

From eqns. (5) and (7),

$$\zeta^*(x) = \mu\nu \begin{cases} (\mu x)^{\delta}, & 0 \leqq x \leqq 1/\mu. \\ (\mu x)^{-\epsilon}, & x \geqq 1/\mu \end{cases} \tag{9}$$

Thus, if our assumptions are correct or approximately so, we should find that the frequency distribution corresponding to $r(s)/E(r|s)$ can be well fit by the distribution given by eqn. (9). Following a suggestion of J. Tukey, we call this the ***double monomial*** distribution. It has two noteworthy features. First, its mode, which occurs at $x = 1/\mu$, is rather sharply peaked, much like the Laplace distribution rather than like the normal distribution. Second, the tail of the double monomial is much higher than is usual in error distributions; it behaves as a negative power of x rather than as an exponential. Thus, very deviant responses will occasionally occur; these make the variance and higher moments relatively unstable quantities and so poor ones to use to estimate parameters.

2. Estimation

Given an empirical frequency distribution corresponding to ζ^*, it is necessary to estimate the parameters δ and ϵ of eqn. (9). The unusual properties of the double monomial distribution, especially its high tail, make suspect some of the more familiar estimation techniques such as equating the observed and theoretical variances. It seems preferable to use some optimal procedure such as maximum likelihood. The distribution we shall fit is somewhat more general than eqn. (9) in that $x_0 = 1/\mu$ is assumed to be a parameter independent of δ and ϵ. Once the estimates of δ, ϵ and x_0 are obtained, we can check to see how well eqn. (8) is satisfied.

Suppose N observations x_i, $i=1, 2, \ldots, N$, are obtained and that they are ordered so that

$$x_1 \leqq x_2 \leqq, \ldots, \leqq x_{N_0} \leqq x_0 \leqq x_{N_0+1} \leqq, \ldots, \leqq x_N.$$

Then if we set equal to zero the partial derivatives with respect to each of the three parameters of the logarithm of the usual likelihood function, we obtain the following simultaneous equations:

$$\begin{aligned} N_0 &= N\left(\frac{\epsilon-1}{\delta+\epsilon}\right), \\ \delta &= \frac{1}{\ln x_0 - \frac{1}{N_0}\sum_{i=1}^{N_0} \ln x_i} - 1, \\ \epsilon &= \frac{1}{\frac{1}{N-N_0}\sum_{i=N_0+1}^{N} \ln x_i - \ln x_0} + 1, \end{aligned} \tag{10}$$

where the logarithms are natural ones. Since explicit solutions have not been obtained, an iteration procedure was programed[1] in Fortran language, and it was used to obtain the estimates given below.

Occasionally the iteration procedure oscillated between two (non-adjacent) integers, in which case we selected an intermediate integer and continued to iterate. In all such cases a unique solution was found.

3. Experimental Work

Experiment 1: Heaviness Judgements

METHOD

Subjects Six female University of Pennsylvania undergraduates were hired on an hourly basis. They were all right-handed and from 19 to 22 years old.

Apparatus To avoid the known weight-size, -colour and -contact area effects (Müller and Schuman, 1889; Payne, 1958, 1961), and to permit the exploration of the whole range of weights that can be comfortably lifted with one finger, we constructed an apparatus that consisted of a wooden box, a lifting arm, a visual shield and a set of 18 weights. The lifting arm passed through a circular hole in the centre of the top of the 10 × 10 × 10 in. box. The arm had a ring at the top, through which S inserted his forefinger, and had a hook at the bottom, to which E attached the weights through the open side of the box. The shield, which was 20 in. high and 14 in. wide, was attached to the top of the box along the open side that faced E; it concealed both E and the weights from S. A light mounted in the shield served as a warning signal.

With no weight attached, the lifter weighed 20 gm. The moulded lead weights, together with the lifter, were spaced approximately logarithmically. Specifically, to an accuracy of about 0·1 gm, they were: 20, 30, 40, 50, 65, 80, 100, 125, 160, 200, 250, 300, 400, 500, 650, 800, 1,000, 1,250 and 1,600 gm. The 20 gm weight, i.e., the lifter alone, was discarded after $S1$ was run because she reported that, on the basis of auditory cues, she could identify it as different from the others.

[1] Our thanks to Mr. Carl Schupp for this program.

Procedure: magnitude estimation Each *S* was run individually in sessions which lasted from one to one and a half hours. Three to four sessions were run per week for a total of between 13 and 15 sessions. Four of these sessions were devoted to the collection of recognition data (see below).

At the start of the first session *S* was told that the purpose of the experiment was to scale her sensation of heaviness of lifted weights. She was instructed to raise the lifter slowly to about 3 cm above the resting position when the warning light flashed, to hold it for about 1 sec, to return it slowly to the resting position, and then to report her numerical judgement of its heaviness as a proportion of a standard weight assigned the value 100. Subject to these limitations, each *S* was permitted to adopt her own mode of lifting, but she was requested not to alter it during the experiment. The instructions were detailed and were amplified in answer to questions or when confusion was apparent. A number of practice trials were given before the experiment to familiarize *S* with the apparatus, procedure, and range of weights.

For *S*s 1, 2 and 3 the standard was the 100 gm weight and for *S*s 4, 5 and 6 it was the 200 gm weight. The standard was presented at the beginning of each session and whenever *S* requested it, which averaged about twice per session. We felt that it would not be desirable to present it more often than this.

The experiment was divided into five blocks, each containing 20 presentations of each stimulus. Within each block the presentations were randomized. The same randomization was used in the first, third and fifth block, and the one that is generated from it by selecting presentations in the order 1, 21, 41, . . ., 2, 22, 42, . . . was used in the second and fourth blocks. A total of about 180 presentations occurred in each ME session. Each stimulus was presented 100 times during the experiment.

The inter-trial interval was approximately 10 sec. Short breaks were permitted when requested; usually there were three or four per session.

Procedure: recognition After each of the first four blocks of the ME experiment we interleaved a session consisting of 100 trials of a recognition experiment. The same apparatus was used, but on each trial one of only two weights, either the standard or the standard plus 20 gm was presented. *S* was instructed to lift exactly as in the ME experiment but to report which of the two weights, the heavier or lighter, was presented. When uncertain, she was instructed to guess. Following each response, *E* informed her whether or not her response had been correct by saying ' right ' or ' wrong '.

EXPERIMENT 2: LOUDNESS JUDGEMENTS

METHOD

Subjects Seven female University of Pennsylvania summer students were hired on an hourly basis. They were from 19 to 29 years old and had no known hearing defects. One *S*, 7, was discarded because she failed to co-operate.

Apparatus A Kohn–Hite model 441-D-1 oscillator fed a 1,000 c.p.s. tone into a Hewlet–Packard model 350 BR attenuator with a shunt resistance of 600 ohms which in turn fed into a filter that gave rise and decay times of 25 msec. The output was then fed into the right earphone of a Grason–Stradler model TDH 39. Calibration of the output was done using an ASA type 1A coupler attached to a Grason–Stradler model 701 condenser microphone and using a Grason–Stradler condenser microphone complement model 726.

Procedure Each *S* was run individually in sessions that lasted about two hours. There were two to three sessions per week for a total of between 8 and 10 sessions, of which four were devoted to recognition trials and the remainder to ME. The overall plan of the experiment was the same as Experiment 1 except that, because of the longer sessions and fewer breaks, approximately 400 presentations occurred per session. The inter-trial interval was again about 10 sec.

The 20 stimuli for the ME experiment were 1,000 c.p.s. tones of 1 sec duration spaced alternatively at 2 and 3 dB from 43 to 90 dB re. 0·0002 dynes/cm^2. For *S*s 8, 9 and

11 the standard was the 70 dB tone and for Ss 10, 12 and 13 it was the 55 dB tone. In the recognition experiment the two stimuli were the 60 dB tone and those shown in column *s* of Table 2.

S was seated in an Industrial Acoustics Corp. sound attenuating chamber. Oral communication was possible via a two-way intercom system. A red warning light occurred 1 sec prior to each stimulus. *S* responded verbally in the ME experiment and by pressing one of two keys in the recognition experiment. As in Experiment 1, feedback was oral in the recognition experiment.

RESULTS

Magnitude Estimation

The arithmetic means of the 100 responses to each of the 18 weights and 20 sound intensities are plotted in double logarithmic coordinates in Figures 1 and 2 respectively. The indicated ' slopes ' are those of visually fitted straight

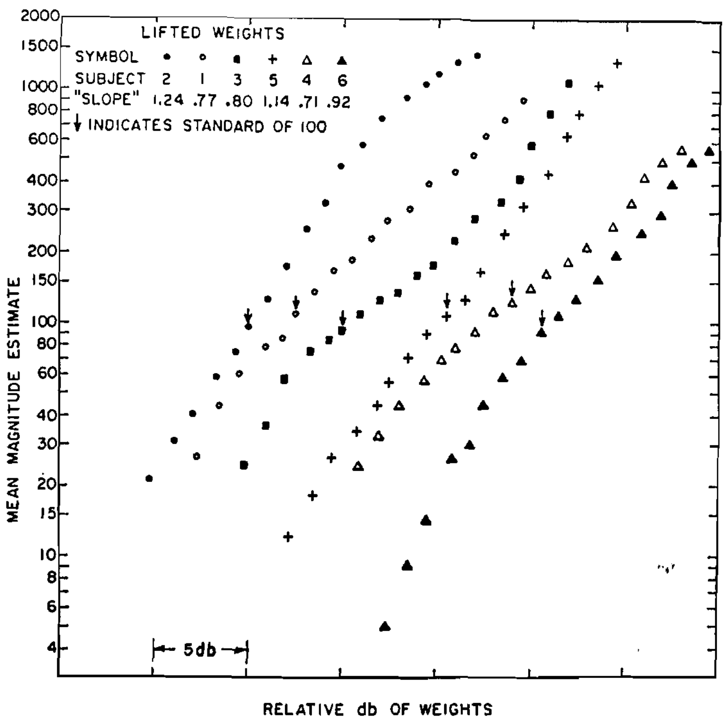

FIGURE 1. Mean magnitude estimates for lifted weights. Each point is the mean of 100 observations. The first point for each *S* is the 30 gm weight and the last is the 1,600 gm one.

lines. In Figure 3 the same data separated by successive fifths i.e., 20 observations per point, are shown for *S*2. The data for other *S*s are similar.

Each response to a stimulus was normalized by dividing it by the mean response to that stimulus, and the resulting distribution over all stimuli was

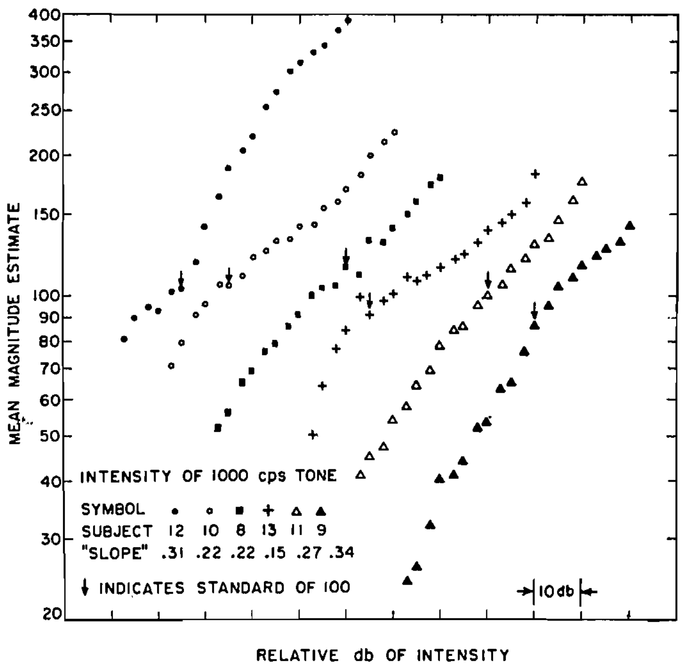

FIGURE 2. Mean magnitude estimates for intensity of 1,000 c.p.s. tone. Each point is the mean of 100 observations. The first point for each *S* is the 43 dB and the last is the 90 dB tone re. 0·0002 dynes/cm.

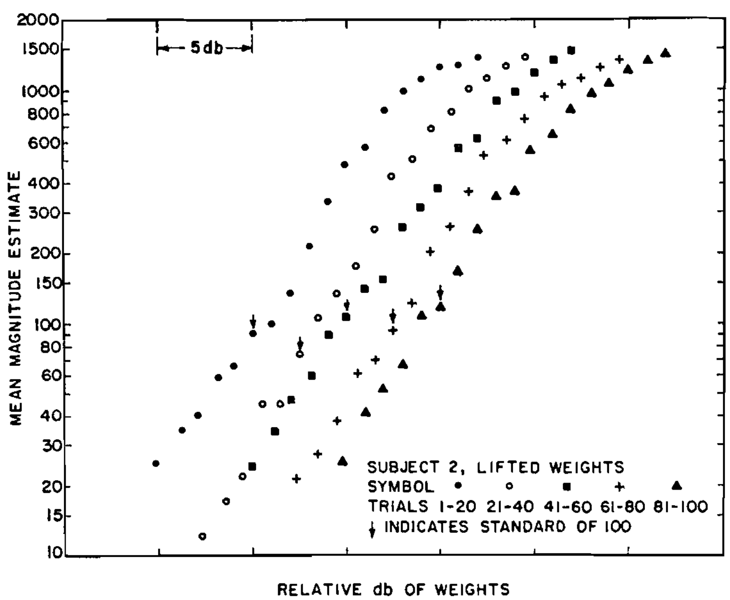

FIGURE 3. Mean magnitude estimates for *S* 2 (lifted weights) grouped by blocks of 20 trials. The first point is, in each case, the 30 gm weight and the last is the 1,600 gm one.

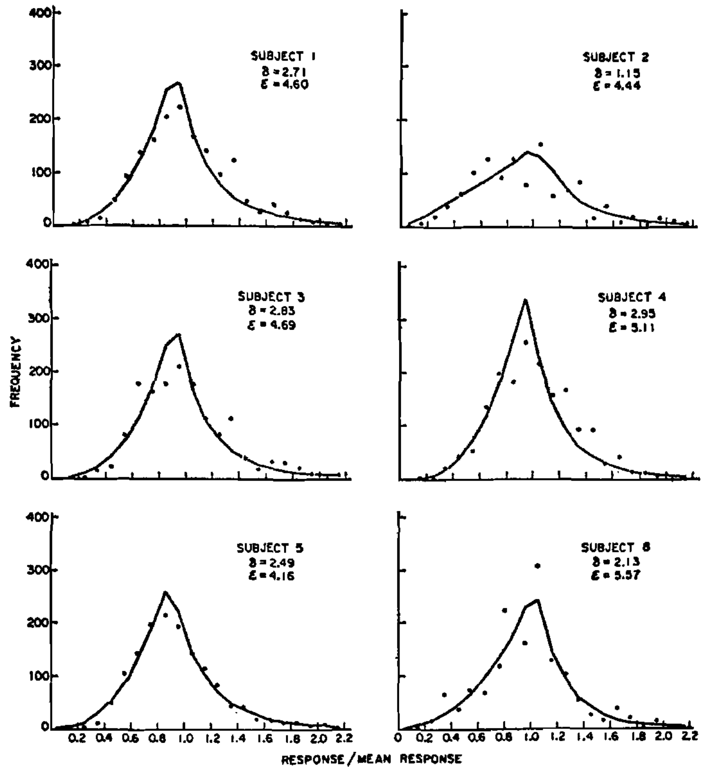

FIGURE 4. Frequency distributions of the ratio of a response to the mean response for the corresponding stimulus for lifted weights. The points are empirical with 0·1 class intervals, and are based on a total of from 1,200 to 1,800 observations in each distribution. The solid lines are the double monomial, with the indicated δ and ϵ estimates, using the same class intervals.

fitted by the double monomial distribution using the iteration scheme previously described. Apparently some *S*s in the lifted weight experiment were able to make absolute identifications of the end stimuli and responded with a very much narrower distribution of numbers; these disparate stimuli have been

TABLE 1. COMPARISON OF x_0 OBTAINED BY ITERATION WITH

$$\frac{1}{\mu}=\left(\frac{\delta+2}{\delta+1}\right)\left(\frac{\epsilon-2}{\epsilon-1}\right)$$

Parameters	*Subjects*											
	1	2	3	4	5	6	8	9	10	11	12	13
x_0	0·92	1·04	0·92	0·95	0·89	1·03	1·03	1·02	1·04	1·00	1·08	1·09
$1/\mu$	0·94	1·05	0·93	0·97	0·89	1·03	1·03	1·03	1·04	1·00	1·07	1·08

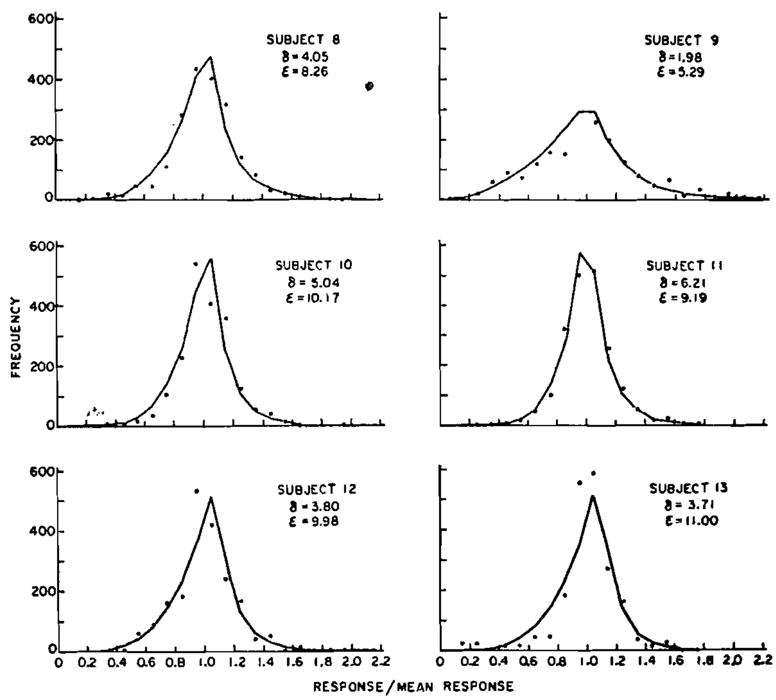

FIGURE 5. Frequency distributions of the ratio of a response to the mean response for the corresponding stimulus for sound intensity. The points are empirical, with 0·1 class intervals, and are based upon a total of 2,000 observations in each distribution. The solid lines are the double monomial, with the indicated δ and ϵ estimates, using the same class intervals.

omitted from the calculations. For the loudness data, the standard deviation of the normalized data decreased by a factor of about 2 over the range used. Because the drop was reasonably regular, no stimuli are omitted. Table 1 compares the value of x_0 obtained from the iteration with the value of $1/\mu$ calculated from eqn. (8) using the iterated estimates of δ and ϵ; it is seen that the two values are substantially the same, thus justifying our use of the more general computer program. The comparisons of the observed and theoretical distributions, using class intervals of 0·1, are shown in Figures 4 and 5.

The distributions of *S*s 8 and 11 were selected as most likely to be fit by the normal distribution in the logarithm of the independent variable. The parameters were estimated using the maximum likelihood method of equating observed and theoretical means and variances. The resulting comparisons are shown in Figure 6.

Recognition

Assuming that the psychophysical function is a power function with slope β, that the distribution of responses found in magnitude estimation is the double

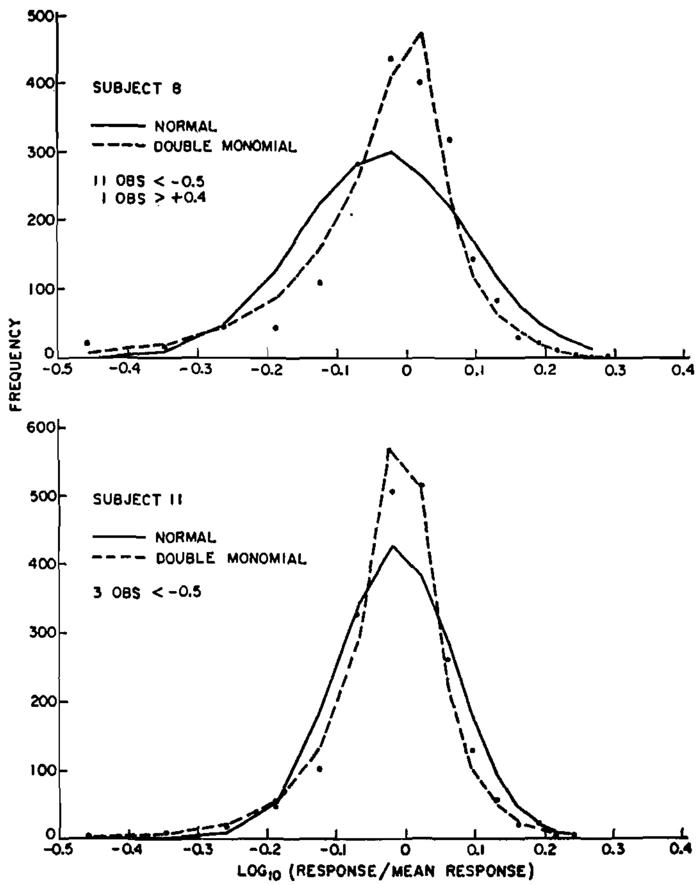

FIGURE 6. Frequency distributions of the logarithm of the ratio of response to a stimulus to the mean response to that stimulus for *S*s 8 and 11. The theoretical curves are the normal and the double monomial. Certain observations are outside the diagram; they are indicated on the plot.

monomial with parameters δ and ϵ, and that the variability of responses in a recognition experiment is due to the same factors as in the magnitude estimation experiment, then Luce and Galanter (1963, p. 293) have shown that

$$p(s|s) = \frac{1}{1+b(s/t)^{\beta\delta}},$$

$$p(s|t) = \frac{1}{1+b(s/t)^{-\beta\epsilon}}, \tag{11}$$

where $s < t$ and b is a bias parameter. Thus, the dB difference corresponding to known $p(s|s)$ and $p(s|t)$ is:

$$\Delta \text{dB} = 10 \log_{10}(t/s) = \frac{10}{\beta(\delta+\epsilon)} \log_{10} \left[\frac{p(s|s)p(t|t)}{p(t|s)p(s|t)} \right]. \quad (12)$$

We can, therefore, compare the actual dB difference used in the recognition experiment with that predicted by eqn. (12) using the values β, δ and ϵ estimated from the ME data and the observed values of $p(s|s)$ and $p(s|t)$. All of these numbers are listed in Table 2.

TABLE 2. COMPARISON OF RECOGNITION DATA AND THEORETICAL PREDICTIONS USING THE MAGNITUDE ESTIMATION DATA

Subject	β	δ	ϵ	s	t	$p(s\|s)$	$p(s\|t)$	ΔdB *actual*	ΔdB *pred.*
1	0·77	2·71	4·60	100	120	0·56	0·34	0·79	0·70
2	1·24	1·15	4·44	100	120	0·90	0·17	0·79	2·37
3	0·80	2·83	4·69	100	120	0·59	0·17	0·79	1·40
4	0·71	2·95	5·11	200	220	0·60	0·30	0·41	0·95
5	1·14	2·49	4·16	200	220	0·67	0·42	0·41	0·58
6	0·92	2·13	5·57	200	220	0·46	0·23	0·41	0·57
8	0·22	4·05	8·25	56	60	0·74	0·14	4·00	4·58
9	0·34	1·98	5·29	56	60	0·82	0·28	4·00	4·30
10	0·22	5·04	10·17	58	60	0·71	0·17	2·00	3·21
11	0·27	6·21	9·19	58	60	0·68	0·24	2·00	1·99
12	0·31	3·80	9·98	56	60	0·83	0·26	4·00	2·69
13	0·15	3·71	11·00	55	60	0·91	0·14	5·00	8·12

Since it is evident from Figures 1 and 2 that power functions, i.e., straight lines in those plots, are not especially good approximations to the data, one can attempt to make better slope estimates in the region of the stimuli used in the recognition experiment. This was done; however, since it altered but little the pattern exhibited in Table 2, we do not report these numbers.

4. DISCUSSION

Form of the Magnitude Estimation Function

No subtle statistical analysis is needed to see that the plots shown in Figures 1 and 2 are not well fitted by straight lines; moreover, even if we introduce the usual ' threshold ' correction on the independent variable, only *S*s 1 and 8 appear to conform to power functions. This conclusion is consistent with the findings of Pradhan and Hoffman (1963), who reported that only one of their six *S*s exhibited a power function for magnitude estimates of heaviness. Their range of ' slopes ' was from 0·45 to 1·91 as compared with our range from 0·71 to 1·24; whether this is a sampling or experimental difference we do not know. Judging by Figure 3 the observed deviations from power functions do not develop slowly over trials; we do not have data adequate to decide whether the deviations develop within the first few trials.

It is not at all apparent from Figures 1 and 2 which, if any, simple family of functions describes these data.

One need not conclude that the underlying psychophysical function fails to be a power function, but only that the mean magnitude estimates for individual *S*s are not power functions. But if we accept that the psychophysical function and the ME scale are not of the same form, then at least one of the assumptions that led to eqn. (3) must be erroneous. Both the postulate that the bias function is a power function, for which there is absolutely no evidence, and that the generalization function depends only upon the ratio of its arguments (eqn. (4)), again for which there is no evidence, are suspect. If either is false, the underlying psychophysical function will, in general, be non-linearly distorted.

To hold to the broad theoretical position in eqn. (1) and, at the same time, to take seriously the suggestion that average group data have the same form as the psychophysical function, namely, a power function, is tenable only if we suppose that these idiosyncratic non-linear distortions average out. No theoretical work has been done to suggest the conditions under which this might reasonably be expected to happen.

Form of the Generalization Function

Although, as we discuss presently, at least one of the assumptions that permitted us to deduce that the generalization function is the double monomial distribution is incorrect, none the less it is evident from Figures 4 and 5 that the average empirical distributions of responses to a stimulus divided by mean response to that stimulus are peculiar in the sense that their modes are sharply peaked, their tails are high, and the distributions are asymmetrical (as indicated by the fact that the estimates of ϵ are from 1·5 to 3·9 times those of δ). The magnitude of these peculiarities is seen (Figure 6) in the failure of the log-normal distribution to fit the empirical data.

Concerning the assumptions that led to the double monomial, the one of no response bias is almost certainly wrong, but how wrong we do not know; and the one that the generalization function depends upon the ratio of its arguments (eqn. (4)) is cast in serious doubt, for at least the loudness data, by the fact that the standard deviation of the observed response distribution normalized by the mean response diminished by a factor of about two over the range of stimulation. Actually, one cannot be certain of the degree to which the failure of the ME function to estimate accurately the psychophysical function is a source of trouble in our attempts to estimate the generalization function. Nevertheless, independent of the assumptions to which we attribute the blame, there can be no doubt that the details of the model are wrong.

The only question that remains about the empirical distributions of responses is whether their unusual qualitative features really are unusual or whether they are artifacts of combining distributions whose standard deviations are as disparate as two to one. Take, for example, the extreme case of two log-normal distributions, the one having a standard deviation twice that of the other, and consider the distribution formed when half the observations are from each. Clearly, such a distribution is both more peaked and has higher tails than a log-normal distribution whose standard deviation equals that of the hybrid. It is not difficult, however, to show by simple calculations that this is not nearly adequate to account for the data and, in any event, it does not explain the observed asymmetry. It appears, therefore, that although we may not know the exact form of the generalization function, not even empirically, it is probably not one of the classical distributions; rather, it is somewhat like the double monomial distribution.

Prediction of the Recognition Data

From Table 2 we see that the prediction of the recognition data from the magnitude estimation data is poor for *S* 2, fair for *S*s 3, 4, 10 and 13, and relatively good for the remaining seven. Our judgements of quality are based on the size of dB differences usually deemed small for these modalities. Since the lore of psychophysics has it that magnitude estimates are a good deal more variable than the responses obtained using classical procedures, it may be surprising that an accurate prediction is possible for over half the *S*s. It is not clear to what extent the failures can be attributed to estimation errors that result from failures in our assumptions. In any event, a reasonable working hypothesis seems to be that much, if not all, of the variability in ME is the same as that exhibited in a recognition experiment run under the same experimental conditions.

5. Conclusions

There is little doubt that these data reject the details of the theory as stated by Luce and Galanter. In our opinion, however, they do not make untenable the general features of that theory, but only some of the specific assumptions. Unfortunately, it is not clear just which ones are wrong. The mean magnitude estimates are, perhaps, sufficiently close to power functions so that, coupled with the large amounts of group data that appear to be power functions, we should retain the hypothesis that psychophysical functions are power functions, but reject the hypothesis that the mean estimate is proportional to the psychophysical function. This last hypothesis can be wrong either because the bias function is not a power function or because the generalization function depends upon both arguments, not just upon their ratio. The failure of either assumption, or both, destroys the argument that led to a double monomial distribution. At the same time, the observed distribution is sufficiently like the double monomial to make one wonder whether relatively slight modifications of the theory would be reasonably adequate. Evidently the theory must be refined, but its main structure does not appear to be completely destroyed by these data.

References

Bush, R. R., Galanter, E. and Luce, R. D. (1963). Characterization and classification of choice experiments. In R. D. Luce, R. R. Bush and E. Galanter (Eds.), *Handbook of mathematical Psychology*, Vol. 1. New York: Wiley. Pp. 77–102.

Dawes, R. M. (1963). The logic of S-R matrices. *Psychol. Rev.* **70**, 365–368.

Luce, R. D. (1963). Detection and recognition. In R. D. Luce, R. R. Bush and E. Galanter (Eds.), *Handbook of Mathematical Psychology*, Vol. 1. New York: Wiley. Pp. 103–189.

Luce, R. D. and Galanter, E. (1963). Psychophysical scaling. In R. D. Luce, R. R. Bush and E. Galanter (Eds.), *Handbook of Mathematical Psychology*, Vol. 1. New York: Wiley. Pp. 245–307.

Müller, G. and Schuman, F. (1889). Über die psychologischen Grundlagen der Vergleichung gehobenen Gewichte. *Arch. ges. Physiol.* **45**, 37–112.

McGill, W. J. (1960). The slope of the loudness function: a puzzle. In H. Gulliksen and S. Messick (Eds.), *Psychological Scaling: Theory and Applications*. New York: Wiley. Pp. 67–81.

Payne, M., Jr. (1958). Apparent weight as a function of colour. *Amer. J. Psychol.* **71**, 724–730.

Payne, M., Jr. (1961). Apparent weight as a function of hue. *Amer. J. Psychol.* **74**, 104–105.

Pradhan, P. L. and Hoffman, P. J. (1963). Effect of spacing and range of stimuli on magnitude estimation judgements. *J. exp. Psychol.* **66**, 533–541.

Rosner, B. G. (1961). Psychophysics and neurophysiology. In S. Koch (Ed.), *Psychology: a Study of a Science*, Vol. 4. New York: McGraw-Hill. Pp. 280–333.

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

Stevens, S. S. (1960). The psychophysics of sensory function. *Am. scient.* **48**, 226–253. Also in W. A. Rosenblith (Ed.), *Sensory Communication*. New York: Wiley (1961). Pp. 1–33.

Stevens, S. S. and Galanter, E. (1957). Ratio scales and category scales for a dozen perceptual continua. *J. exp. Psychol.* **54**, 377–411.