An experimental comparison of stimulus and response translated power functions for brightness¹

ROBERT F. FAGOT AND MANARD R. STEWART

UNIVERSITY OF OREGON

A threshold parameter has been introduced into the power law in two different ways, one involving translation on the stimulus axis (the Φ -Law) and the second involving translation on the response axis (the Ψ -Law). In this study the two laws were compared using the data from 6 Ss making magnitude estimation judgments of brightness. Although the prevalent view appears to favor the Φ -Law, the data of this study clearly dictate a choice of the Ψ -Law over the Φ -Law for brightness.

Marks and J. C. Stevens (1966) have claimed that the relation between brightness (Ψ) and luminance (Φ) has been shown to have the form

$$\Psi = a(\Phi - t)^k, \quad (1)$$

where t is a threshold parameter. Data in support of this form of the power law for such diverse attributes as brightness, loudness, hardness and softness, apparent force, vibration, and electric shock have been provided by Aiba and Stevens (1964), Harper and Stevens (1964), Schneider and Lane (1963), J. C. Stevens and Guirao (1964), J. C. Stevens and Mack (1959), J. C. Stevens, Mack, and S. S. Stevens (1960), Stevens (1956a, b, 1959, 1966a, b), Stevens and Guirao (1963), and Stevens and Harris (1962). This form of the law has been called the *Stimulus Law* by Fagot (1966), since it effects translation on the intensity axis, and in this paper will be designated the Φ -Law. It is clear that the effect of the translation is to measure intensity from the threshold rather than from the origin.

In spite of the apparent abundance of evidence favoring this form of the Φ -Law, there is at least one alternative which should be given serious consideration, as suggested first by McGill's (1960) data, from which it was concluded that a power function with additive constant

$$\Psi = a\Phi^k + b, \quad (2)$$

was the rule for individual Os making magnitude judgments of loudness. If a threshold parameter t is introduced into Equation (2) with the constraint that the function must pass through the point $(\Phi, \Psi) = (t, 0)$, then Equation (3)

$$\Psi = a(\Phi^k - t^k) \quad (3)$$

is the mathematical consequence. This form of the law has been designated the *Response Law*, since translation is effected on the psychological axis, and henceforth shall be referred to as the Ψ -Law.

Zwislocki and Hellman (1960) concluded that the Ψ -Law provided a better fit to the data of various sense modalities than the Φ -Law. Hellman and Zwislocki (1961) investigated the form of the loudness function near threshold by the method of magnitude estimation, and Lochner and Burger (1961) carried out an analysis of the Hellman and Zwislocki data from which they concluded that the Ψ -Law provided a better fit than the Φ -Law. Finally, Stewart, Fagot, and Eskildsen (1967) found that for fractionation judgments of brightness the Ψ -Law provided a better fit to the data, although neither form of the law fit well.

In addition to these conflicting studies, methodological considerations suggest that tacit acceptance of the Φ -Law is premature. First, none of the studies purporting to support the Φ -Law used statistical procedures to determine the best-fitting theoretical power function, and conclusions of "good fit" rested on inspection of curves drawn by hand. Second, and of greater importance, in none of these studies are plausible alternatives, e.g., the Ψ -Law, compared to the Φ -Law. (In fact, in only one of these studies (Marks & J. C. Stevens, 1966) is the Ψ -Law mentioned as an alternative.) Hence, although the prevalent view appears to be that the Φ -Law is the appropriate form of the power law, adequate comparative studies are lacking. The purpose of the experiment to be outlined is to choose between the two forms of the power law based on a comparison of the best-fitting individual brightness power functions obtained from magnitude estimation data.

METHOD

Subjects

Subjects were three male graduate students in psychology, and three females, the latter being wives of university students. Four of the Ss had participated in a single magnitude estimation session prior to the experiment; the remaining two Ss had had no previous experience with the magnitude estimation task.

Apparatus

The apparatus, described in detail elsewhere (Es-

Table 1. Estimates of Exponents (k) and Thresholds (t) for Various Forms of the Power Function and Interpretations of t

Subjects	Interpretation of t					
	Absolute Threshold		Effective Threshold		Indirect Threshold	
	Φ Law	Ψ Law	Φ Law	Ψ Law	Φ Law	Ψ Law
G. C. k	.54	.54	.53	.50	.50	.47
t		.25		9.0	35.67	19.13
D. C. k	.74	.73	.72	.68	.61	.53
t		.36		21.0	49.88	48.69
D. A. k	.45	.44	.45	.39	.41	.34
t		.48		9.9	46.58	24.77
S. A. k	.72	.72	.72	.70	.65	.59
t		1.14		8.4	48.29	42.14
J. K. k	.29	.25	.29	.17	.25	.07
t		.60		11.0	48.86	29.67
A. W. k	.29	.25	.29	.18	.26	.13
t		.55		9.3	48.29	18.22

Note: Threshold entries (t) have been multiplied by 10^{-4} for tabular presentation.

kildsen, 1963), provided two luminous white circular targets 13 mm in diameter, placed horizontally and separated by 90 mm, center to center. The target to S's right was held at a constant luminance—10 ft-L—throughout the entire experiment and was designated the standard. The target to S's left was under the control of E and on each trial adjusted by E to one of the 11 luminances chosen for the experiment. These luminances Φ_1 through Φ_{11} , were .005, .010, .020, .040, .080, .150, .300, .650, 1.30, 2.50, and 5.00 ft-L, selected according to two principles: (a) adjacent ratios in the series should be approximately 1/2; and (b) stimuli should be selected close to the threshold where discrimination between alternative forms of the power law is maximal (Fagot, 1966).

Procedure

Subjects participated in pairs for a total of five sessions, each lasting about 1 h and separated by at least 4 h. For practical reasons, it was necessary to space the five sessions over four days, hence on one of the days two sessions were required. Within each session S made 77 magnitude estimation judgments (relative to standard), seven for each of the 11 stimuli, which is a total of 35 estimations of each comparison stimulus, and a grand total of

385 estimations for each S. The order of presentation of the stimuli consisted of random permutations of the 11 stimuli, with the restriction that the last stimulus of permutation $n-1$ could not be the first stimulus of permutation n .

The final session was lengthened to obtain measures of the "absolute" and "effective" thresholds. Ten observations in the absence of a masking stimulus were used to estimate the absolute threshold, and 10 observations in the presence of the standard were used to estimate the effective threshold, since the standard was the largest stimulus used in the experiment and was present for all observations. The standard stimulus was selected as the largest stimulus in order that the effective threshold would be constant for all observations. The method of minimal change (Guilford, 1954, pp. 101-107) was used for both thresholds.

Prior to the first session S was given instructions concerning the task. The experimental room was illuminated during the instructions, and the pair of Ss were given a few practice trials. The following set of instructions were read to each pair of Ss:

"On each trial you will be presented a pair of lights. The light on your right will be the same brightness throughout the entire experiment. We shall call this light the standard, and you will be asked to judge the light on your left in relation to the standard. The brightness of the standard will be designated 1000, and on each trial your task will be to assign a number to the other light in such a way as to reflect its apparent brightness. For example, if the light on your left appears to be half as bright as the standard, you should record your estimate as 500. Use any positive number you feel represents the magnitude of the light on your left in relation to the standard of 1000.

"The beginning of a trial will be signaled by the appearance of a pair of lights. As soon as you have decided upon the appropriate value for the stimulus, press the button on the forward part of your chair (E indicates the position of the button by pointing). Be sure to hold the button down until the shutter closes over the lights. After the shutter closes you are to record your response; in order to see to do this, press the lower button controlling the lamp attached to the desk (again E indicates the position of the button by pointing). The lamp should be used only when recording your response. When you have recorded your estimate, push the button on the forward part of your chair again. Hold the button down until you hear a click. This will signal the experimenter you are ready for the next trial.

"We would like to caution you against basing your judgment on qualities of the light other than its apparent brightness. For example, to some people dim lights appear to be farther away, and they could base their judgment of brightness on the distance

Table 2. Relative Thresholds

Subject	t_E/t_A	t_1/t_E	
		Φ Law	Ψ Law
G. C.	36.00	3.96	2.13
D. C.	58.33	2.38	2.32
D. A.	20.63	4.71	2.50
S. A.	7.37	5.75	5.02
J. K.	18.33	4.44	2.70
A. W.	16.91	5.19	1.96

of the light. Do not use cues of this sort in making your judgments of brightness, use only the apparent brightness of the light.

"Are there any questions? Now let's run through the procedure a few times before starting the experiment."

Before each experimental session Ss were dark adapted for 12-15 min in the light-tight experimental room. The S's use of the lamp was restricted to the time he used in recording, and the light was a dim red neon globe providing an indirect light for S to record. At the brightest point on a white sheet placed below the bulb the intensity was approximately .25 ft-L, and should have had no effect on the threshold.

During the final session E was required to place a filter over the target between the magnitude estimation task and the threshold measurement task. During the time E changed filters, S was asked to sit with his eyes closed.

RESULTS

In analyzing the data we considered three possible interpretations of t in Equations (1) and (2). Case A: t was interpreted as the absolute threshold (t_A) and measured by the method of minimal change in the absence of a masking stimulus. Case E: t was

interpreted as the effective threshold (t_E) and measured by the method of minimal change in the presence of the standard. Since the standard was the highest intensity stimulus presented to S, it was assumed that t_E was constant over trials in which different comparison stimuli were presented in conjunction with the standard. This assumption is quite reasonable in this case, but would be dubious if the standard was lower in intensity than some of the comparison stimuli. Case I: t was interpreted as the indirect threshold (t_I), and in this case, t_I and k were jointly estimated from the magnitude estimation data. In Cases A and E, t_A and t_E were independently measured and only k was estimated from the magnitude estimation data. In particular, the function

$$T = \sum_{i=1}^{11} [\bar{Y}_i - \log f(\Phi_i)]^2 \quad (4)$$

was minimized by numerical methods using an IBM 360 (50) computer.² In Equation (4),

$$\bar{Y}_i = (1/35) \sum_{j=1}^{35} \log \Psi_{ij},$$

Fig. 1a. Plots of $\log \Psi$ as a function of $\log \Phi$ for t interpreted as the indirect threshold with the Φ -Law plotted on the left and the Ψ -Law on the right (Ss DA, DC, and GC).

Fig. 1b. Plots of $\log \Psi$ as a function of $\log \Phi$ for t interpreted as the indirect threshold with the Φ -Law plotted on the left and the Ψ -Law on the right (Ss AW, JK, and SA).

where Ψ_{ij} is the j th magnitude estimation of comparison stimulus Φ_i , and f represents the theoretical function (the particular form of the power law). Thus for Cases A and E the estimate of k was chosen which minimized T , and for Case I joint estimates of k and t were chosen which minimized T . For all three cases the theoretical function was constrained to pass through the point $(\Phi, \Psi) = (10, 1000)$, since Ss were instructed to assign 1000 to the standard stimulus of 10 ft-L. In addition, for Cases A and E, the necessary constraint was imposed that the theoretical function must pass through the point $(\Phi, \Psi) = (t, 0)$, where, of course, t is independently measured as t_A and t_E for Cases A and E, respectively. A separate analysis was carried out for each of the six Ss, and for each of the three possible interpretations of t .

$\log \Psi$, rather than Ψ , was used as the dependent variable in order to obtain homogeneity of variance, since homoscedasticity is assumed in the regression analyses which follow. Bartlett's test (Winer, 1962, p. 95) was applied and homogeneity of variance ($\log \Psi$) was rejected for only one (DC) of the six Ss.

The estimates of t and k for each S, each law, and each interpretation of t are given in Table 1.

Table 2 presents relative thresholds, i.e., the ratios of effective threshold to absolute threshold (t_E/t_A) and the ratios of indirect threshold to effective threshold (t_I/t_E) for each law. Note that the indirect thresholds are considerably larger than the effective thresholds for both laws, but that the Ψ -Law estimates of t_I are much closer to t_E for every S.

Figures 1a and 1b exhibit the fits (for each S) of each form of the power function for Case I. The solid line is a theoretical function showing $\log \Psi$ as a function of $\log \Phi$, whereas the solid circles are data points. The log-log plots are not, of course, straight lines as should be the case for the simple power function ($\Psi = a\Phi^k$). Casual inspection of the graphs suggests a rather good fit for three of the Ss (GC, DA, and AW), especially for the Ψ -Law, but systematic deviations from both forms of the law for the remaining three Ss.

Table 3 summarizes the significance tests resulting from the regression analysis performed separately for each of the two forms of the psychophysical function under the conditions of Case I. Inspection of this table shows that the effect of deviation from the theoretical Φ -Law was significant for each of the six Ss ($p < .001$ for five Ss and $p <$

Fig. 2a. Plots of $\log \Psi$ as a function of $\log \Phi$ for t interpreted as the effective threshold with the Φ -Law plotted on the left and the Ψ -Law on the right (Ss DA, DC, and GC).

Fig. 2b. Plots of $\log \Psi$ as a function of $\log \Phi$ for t interpreted as the effective threshold with the Φ -Law plotted on the left and the Ψ -Law on the right (Ss AW, JK, and SA).

Table 3.
F Values^a for Regression Analysis of Variance with $t = t_1$

Subject	Φ Law Deviation	Ψ Law Deviation
G. C.	4.97**	1.99*
D. C.	19.51**	13.44**
D. A.	2.66**	2.63**
S. A.	41.48**	27.30**
J. K.	22.97**	7.48**
A. W.	5.10**	1.02

^a Based on 8 and 374 df
* $p < .05$, ** $p < .01$

.01 for one S). The effect of deviation from the theoretical Ψ -Law was significant for five of the six Ss, with $p < .05$ in one of the five cases. For each of the six Ss, the F ratio was larger in the case of the Φ -Law. Of the three Ss (GC, DA, and AW) who appeared to fit well from inspection of the graphs in Figs. 1a and 1b, only in the case of AW for the Ψ -Law was the effect of deviations not significant, although the F in the case of GC was barely significant at the .05 level for the Ψ -Law.

Figures 2a and 2b exhibit the fits of each form of the power function for Case E. As in the case of Figs. 1a and 1b, the graphs of Ss GC, DA, and AW appear to give good fits to both forms of the law, although perceivably better for the Ψ -Law. However, as Table 4 shows, the effects of deviations were significant for all Ss for both forms of the law. However, as in the case of the t_1 interpretation, for all Ss, the F ratio was smaller in the case of the Ψ -Law. The graphs in Figs. 2a and 2b compare surprisingly favorably with those of Figs. 1a and 1b in view of the fact that the effective threshold is independently measured and theoretically cannot provide a better fit than under the t_1 interpretation.

Table 5 gives the results of the regression analysis under Condition A. With t interpreted as the absolute threshold, the effect of deviation from the theoretical power curve was significant ($p < .001$) for all six Ss, under both forms of the power law, and for this reason plots under this condition are omitted.

One way to compare the two laws based on the regression analyses would be in terms of the relative deviation mean squares, which essentially give the estimated variance of errors of prediction. The

Table 4.
F Values^a for Regression Analysis of Variance with $t = t_E$

Subject	Φ Law Deviation	Ψ Law Deviation
G. C.	10.50**	4.22**
D. C.	48.08**	41.56**
D. A.	24.00**	9.83**
S. A.	75.18**	66.73**
J. K.	44.01**	18.53**
A. W.	13.09**	2.65**

^a Based on 9 and 374 df
** $p < .01$

Table 5.
F Values^a for Regression Analysis of Variance with $t = t_A$

Subject	Φ Law Deviation	Ψ Law Deviation
G. C.	12.95**	11.78**
D. C.	55.62**	55.22**
D. A.	28.29**	24.35**
S. A.	80.58**	78.28**
J. K.	47.22**	38.21**
A. W.	14.55**	10.17**

^a Based on 9 and 374 df
** $p < .01$

ratios of the deviation mean squares are given in Table 6, where $\hat{\sigma}_e^2(\Phi)$ and $\hat{\sigma}_e^2(\Psi)$ denote the deviation mean squares for the Φ -Law and Ψ -Law, respectively. Note that all entries in the table are greater than one, indicating that for all six Ss under all three interpretations of t , the variance of the errors of prediction was larger in the case of the Φ -Law.³ Under the t_1 interpretation, except for DA for whom the mean squares are almost equal, the Ψ -Law shows a marked superiority, doing best with AW whose Φ -Law deviation mean square is about five times as large as for the Ψ -Law. The comparative situation under the t_E situation does not differ appreciably from that of t_1 , although the pattern over Ss differs somewhat. There is, however, a marked difference under the t_A interpretation; here we see that all ratios are much closer to unity, the largest being 1.43. The reason for this is quite obvious; the smaller the threshold parameter, the smaller the difference between the two forms of the power law. In the limit, when $t = 0$, both forms are exactly equivalent.

An alternative method of comparing the errors of prediction for the two laws is to focus on the absolute errors made by each law relative to each S and each data point, and then compare the two by means of a simple sign test. This analysis follows:

Let $d_{ij}(\Psi)$ denote the absolute deviation between the scale value predicted by the Ψ -Law and the observed magnitude estimation (the Ψ -Law absolute error) for the i th S and j th comparison stimulus, and $d_{ij}(\Phi)$ the corresponding absolute deviation for the Φ -Law. Define

$$D_{ij} = 1 \text{ if } d_{ij}(\Psi) < d_{ij}(\Phi) \quad j = 1, \dots, 11 \quad (\text{stimuli})$$

$$D_{ij} = 0 \text{ if } d_{ij}(\Psi) \geq d_{ij}(\Phi) \quad i = 1, \dots, 6 \quad (\text{Ss}).$$

Then

$$\hat{p} = (1/66) \sum_i \sum_j D_{ij} \quad (5)$$

gives the proportion of times the Ψ -Law predictions were closer to the data points (or the errors smaller) than the Φ -Law predictions. If the two laws are equally good descriptions of magnitude estima-

Table 6. Ratios of Deviation Mean Squares

Subjects	$\frac{\sigma_e^2(\Phi)}{\sigma_e^2(\Psi)}$		
	t_I	t_E	t_A
G. C.	2.50	2.49	1.10
D. C.	1.45	1.16	1.01
D. A.	1.01	2.44	1.16
S. A.	1.52	1.13	1.03
J. K.	3.07	2.38	1.24
A. W.	5.01	4.94	1.43

tion data, then $p=1/2$, which is the hypothesis we wish to test. Set $\alpha = .01$ for a two-tail test; then since $\sqrt{PQ/N} = \sqrt{(.50)(.50)/66}$ and $Z_{.995} = 2.58$, the decision procedure will be as follows: accept the Ψ -Law if $\hat{p} \geq .659$; accept the Φ -Law if $\hat{p} \leq .341$; and reserve judgment if $.341 < \hat{p} < .659$.

This test was carried out for each of the three interpretations of t . Denote by $\hat{p}(t_I)$ the proportion of times the error was smaller for the Ψ -Law given that t is the indirect threshold, and $\hat{p}(t_A)$ and $\hat{p}(t_E)$ corresponding estimates for Cases A and E, respectively. Then the data were:

$$\hat{p}(t_I) = 49/66 = .74$$

$$\hat{p}(t_A) = \hat{p}(t_E) = 62/66 = .94$$

And, hence, for all three cases the Ψ -Law is accepted and the Φ -Law rejected. In fact, α could have been set as low as $\alpha = .0001$ and the Ψ -Law would still have been accepted over the Φ -Law, which is overwhelming evidence in favor of the Ψ -Law.

A final analysis of the data follows a method outlined by Fagot (1966) and applied by Stewart, Fagot, and Eskildsen (1967), using the exponents as the dependent variable. The method as applied to magnitude estimation involves the estimation of k relative to judgments of a single comparison stimulus, given an independent measure of t . The rationale is that if a particular form of the power law holds, then estimates of the exponent of that form of the law should be independent of the stimuli used to make the estimate. In order to test this hypothesis of invariance of exponents, a three-way (stimulus by day by S) factorial design (treating stimuli and days as fixed factors and Ss as a random factor) was carried out separately for each form of the law and each of the two independent measures of t — t_A and t_E . For both independent measures of t the result was the same: the effect of stimuli was significant at the .01 level for the Φ -Law, but only at the .05 level for the Ψ -Law. The effect of interaction between Ss and stimuli was also significant ($p < .001$), as would be expected from the previous analysis which partitioned the six Ss into two groups, one of which showed markedly poorer fit. The sig-

nificant main effect and interaction effect at the same levels of significance for both laws was precisely the result obtained by Stewart, Fagot, and Eskildsen (1967) for the effect of standard on fractionation judgments of brightness.

DISCUSSION

The analysis of the errors of prediction for the two laws by means of a sign test led to a result that, on the surface, might appear to be in conflict with the results given in Table 6 giving the relative deviation mean squares for the two laws. In the latter it was found that the ratio of deviation mean squares was much closer to unity when t is interpreted as the absolute threshold, yet the proportion of times the absolute error was smaller for the Ψ -Law was much larger (94%) for the t_A and t_E interpretations than for the t_I interpretation (74%). In the case of the t_A and t_E interpretations, both functions are fixed at both the upper and lower ends, and intersect at one intermediate point such that the Ψ -function lies above the Φ -function for all points above the point of intersection and below the Φ -function for all points below the point of intersection. The data points, however, tend to be mostly above both curves above the point of intersection and below both curves below the point of intersection. Hence, although the two functions tend to converge as t becomes small, the data points are consistently closer to the Ψ -function for $t \neq 0$. With the threshold treated as a free parameter (t_I), the estimates of t differ for the two laws and are much larger in both cases, consequently the two functions diverge more and data points tend more to lie between the two functions, and hence, through the operation of random error, a somewhat higher proportion of the data points will lie closer to the Φ -Law. However, the ratio of deviation mean squares, due to the larger t and greater discrepancy between the two functions, will increase substantially. Hence, even when t is small and the two functions close together, the data permit a choice of the Ψ -Law over the Φ -Law.

Another point to be noted in this context is the relative magnitudes of t_I for each of the two laws. Table 2 indicates that t_I averages (over the six Ss) 4.4 times as large as t_E for the Φ -Law and only 2.8 for the Ψ -Law. The significance of this lies in the fact that other possible independent measures of the threshold may give values higher than t_E but almost certainly less than either t_I . For this reason, it is highly likely that all independent measures of t will result in a better fit for the Ψ -Law. This is an interesting likelihood since the psychophysical function would be a much more impressive "theory" if good fits could be obtained with an independently measured parameter t .

Table 1 indicates that the exponents for these

Ss are somewhat higher than usually found. However, if we discard the three Ss who showed very poor fits, the values are not too discrepant: .50, .41, and .26 (mean .39) for the Φ -Law and .47, .34, and .13 (mean .31) for the Ψ -Law (the Ψ -Law exponents will always be lower than the Φ -Law exponents based on the same data), whereas Stevens and Galanter (1957) report a group exponent for brightness around .30 for the simple power law. The difference may be due to the lower range of stimuli used in this experiment, consistent with the result obtained by Stewart, Fagot, and Eskildsen (1967) which showed that exponents (obtained from fractionation judgments) tended to decrease with increases in the intensity of the stimuli.

Inspection of the graphs in Figs. 1 and 2 raises problems of interpretation. First of all, the graphs for the three "poor" Ss (DC, SA, JK) all exhibit marked systematic deviations: it appears that all three Ss tended to overestimate high intensity stimuli relative to low intensity stimuli, based on theoretical (power law) expectations. It should be recalled at this point that the power functions were constrained to pass through the point (10, 1000), since S was instructed to assign 1000 to the standard stimulus of 10 ft-L, and judge the comparison stimulus in relation to the value of 1000 for the standard. Inspection of the graphs suggests the possibility that if the functions were not constrained to pass through this point, the fits would be better and the systematic deviations perhaps eliminated. Such an analysis would introduce an additional parameter (a of Equations (1) and (3)) and *a priori* must fit at least as well as the analysis actually carried out. However, such an analysis, involving the estimation of the psychological magnitude of the standard, would make little sense in view of the fact that S's magnitude estimations were made relative to the assignment of 1000 to the standard. Furthermore, such an analysis could obscure possibly real systematic deviations, which ought to be explored. In this connection, Marks and Stevens (1966) tend to play down the salience of both random⁴ and systematic deviations from the power function (Φ -Law). However, in view of the "visual" methods of curve fitting employed in that study, it is difficult to see how any safe conclusions could be drawn regarding the presence or absence of systematic deviations.

In attempting to explain the apparent tendency of these three Ss to overestimate high intensity stimuli (below the standard) and underestimate low intensity stimuli, we considered that a test of the psychophysical law is in fact confounded with the implicit hypothesis of the validity of the magnitude estimation scale. For this reason we examined the estimation behavior of our Ss in the following way.

For each S we computed the mean number of

numerical assignments used per day, as well as the total number used over the five day period. The mean number of assignments used per day was 23.1 for the three poor Ss and 35.2 for the three good Ss, with no overlap; i.e., the three good Ss used about 50% more assignments. The total number of assignments used averaged 38.0 for the three poor Ss and 60.3 for the three good Ss, which is almost 60% more assignments for the good Ss. The latter figure indicates that the poor Ss used about 3.5 assignments per stimulus, and the good Ss about 5.5.

The finding that poor Ss used markedly fewer numerical assignments could be interpreted as implying that these Ss tended to make "category" judgments rather than magnitude estimations; i.e., to adopt at the outset the strategy of using a limited number of assignments or categories and on each trial try to "identify" the stimulus, i.e., assign it to its proper category. This interpretation seems quite reasonable for JK, who used the smallest number of numerical assignments, and for the first three days used only the assignments 100, 200, ..., 1000 plus 50. Since the stimuli were logarithmically spaced (in ft-L) and this S set up equally spaced numerical assignments (except for the assignment 50), his psychophysical function would have been approximately logarithmic if he had properly classified all stimuli according to this strategy. In spite of the fact that this S made many errors of classification in relation to this inferred category strategy (and apparently abandoned the strategy on the fourth and fifth days), it is interesting that his exponent was in fact .07 for the Ψ -Law under the t_1 interpretation, not far from a log law, and closer to the magnitude exponent usually obtained for category scaling rather than magnitude estimation scaling. However, the other two poor Ss actually had the highest exponents among the six Ss. Nevertheless, the number of numerical assignments used sharply discriminates between these two groups, and suggests the possibility that the poor Ss were not adequately carrying out the assigned magnitude estimation task. The result argues for separating the two confounded hypotheses, namely that of the psychophysical law and the validity of the magnitude estimation scale, by first carrying out tests of scale internal consistency.

We may also note that the apparent tendency of the three poor Ss to overestimate high stimuli may have nothing to do with the location of the standard at the high end of the scale, but (to the extent that the category strategy hypothesis is tenable) may be brought about by a tendency to broaden the categories at the upper end.

Focusing now on the good Ss (GC, DA, and AW), for whom there are no apparent systematic deviations, we are still faced with a problem of inter-

pretation, since visual inspection of the graphs suggests a good fit for both laws, yet the fit of all three Ss is rejected by the Φ -Law, and only one (AW) is accepted by the Ψ -Law. This means that as small as the deviations mean square may appear to be from inspection of the graphs, it is large relative to the very small error mean square for each of the three good Ss. Fortunately, it is not the case that the smaller F ratios for these three Ss are purchased at the expense of larger error mean squares than for the three poor Ss. The comparative error mean squares (in log units) are .10, .14, and .15 for the three good Ss, and .13, .24, and 1.07 for the three poor Ss. The very great power of these tests is obtained through the use of very large degrees of freedom—374—for the error term. To illustrate the power considerations involved here, consider the following alternative analysis of the data. Each comparison stimulus was evaluated seven times in each session. Instead of treating the seven magnitude estimations as seven observations, we took the mean of these seven observations, and treated this mean as one observation, which gave a total of 55 observations, rather than the 385 of the original analysis. It might be argued that such a procedure is very reasonable, since these means are more likely to be normal and independent. In any event, this reduces the degrees of freedom in the error term to 44, and results in an increase in the size of the error mean square, whereas, of course, the deviations mean square does not change. With this analysis, the effect of deviations from the Φ -Law ($t = t_p$) was not significant for one of the three good Ss, and the effect of deviations from the Ψ -Law was not significant for any of the three good Ss. The effects of deviations for the three poor Ss were still significant for both laws, although the F ratios were in all cases smaller. The point is that due to power considerations models should not ordinarily be evaluated in "isolation" by a goodness-of-fit test, since any model is at best an approximation, and hence can always be rejected by a sufficiently powerful goodness-of-fit test. A more reasonable procedure, when feasible, is to compare models, and choose the model which gives the best account of the data.

How do the two laws fare judged against the same data? As mentioned previously, the prevalent view appears to favor the Φ -Law. However, the visual methods of curve fitting employed in many of the studies supporting the Φ -Law make it difficult to take seriously the assertions that the Φ -Law provides good fits to the data. The main difficulty, however, is that the studies purporting to support the Φ -Law did not compare the two laws based on the data.

It is rather surprising that some recent writers (e.g., Marks & J. C. Stevens, 1966) take the Φ -Law

for granted, in view of previous contradictory evidence (e.g., McGill, 1960; Zwischlocki & Hellman, 1960; Lochner & Burger, 1961). The recent study previously cited by Stewart, Fagot, and Eskildsen (1967), in which they studied the effect of a varying standard on fractionation judgments of brightness, also favored the Ψ -Law. In the latter study, for each of five Ss, the variance due to standard was greater for the Φ -Law.

All analyses made in the present investigation favor the Ψ -Law. Especially convincing is the overall comparison of "errors" (deviation of the theoretical prediction for a particular law from the data point) which showed that when t is interpreted as the indirect threshold, the Ψ -Law made smaller errors in 74% of the cases ($p < .00006$), and when t is interpreted as the absolute or effective threshold, the Ψ -Law made smaller errors in 94% of the cases. The data are sharply critical of the prevalent view that the Φ -Law is the appropriate form of the power law, and given a choice between the two, these data unequivocally dictate a choice of the Ψ -Law for brightness. It would be of considerable interest to determine if such a choice could also be made for other attributes, such as loudness, heaviness, warmth, etc., and other methods, such as fractionation. The psychophysical law as a theoretical system would be much less satisfying if the particular form of the law depended upon the attribute scaled, or possibly the method used.

What are the psychological implications of the conclusion that the Ψ -Law is the appropriate form of the power law for brightness? Following the rationalization given by Lochner and Burger (1961) for loudness, we would assume that the fundamental relation between luminance and brightness is given by a simple power law, and that the reduction in brightness to zero at the threshold is due to masking by physiological noise. This *constant brightness* is then subtracted from the brightness of a stimulus at all levels, according to the Ψ -Law. The Φ -Law, on the other hand, implies an effective reduction of *luminance* by a constant amount, and hence accepting the Ψ -Law for brightness implies an effective reduction in brightness rather than luminance, which seems quite reasonable. However, unless some method (other than the method of minimal change) produces a threshold closer to the indirect threshold measured from the magnitude estimation data, some doubt is cast on the interpretation of t as a threshold.

References

- Aiba, T. S., & Stevens, S. S. Relation of brightness to duration and luminance under light-and-dark-adaptation. *Vis. Res.*, 1964, 4, 391-401.
- Eskildsen, P. R. An apparatus for studies of brightness which controls variation of intensity with time. *Amer. J. Psychol.*, 1963, 76, 321-323.
- Fagot, R. F. Alternative power laws for ratio scaling. *Psychometrika*, 1966, 31, 201-214.

- Guilford, J. P. *Psychometric methods*. New York: McGraw-Hill, 1954.
- Harper, R., & Stevens, S. S. Subjective hardness of compliant materials. *Quart. J. exp. Psychol.*, 1964, 16, 204-215.
- Hellman, R. P., & Zwislowski, J. Some factors affecting the estimation of loudness. *J. Acoust. Soc. Amer.*, 1961, 33, 689-694.
- Lochner, J. P. A., & Burger, J. F. Form of the loudness function in the presence of masking noise. *J. Acoust. Soc. Amer.*, 1961, 33, 1705-1707.
- McGill, W. A. The slope of the loudness function: A puzzle. In H. Gulliksen & S. Messick (Eds.), *Psychological scaling: theory and application*. New York: Wiley, 1960.
- Marks, L. E., & Stevens, J. C. Individual brightness functions. *Percept. & Psychophys.*, 1966, 1, 17-24.
- Schneider, B., & Lane, H. Ratio scales, category scales and variability in the production of loudness and softness. *J. Acoust. Soc. Amer.*, 1963, 35, 1953-1961.
- Stevens, J. C., & Guirao, Miguelina. Individual loudness functions. *J. Acoust. Soc. Amer.*, 1964, 36, 2210-2213.
- Stevens, J. C., & Mack, J. D. Scales of apparent force. *J. exp. Psychol.*, 1959, 58, 405-413.
- Stevens, J. C., Mack, J. D., & Stevens, S. S. Growth of sensation on seven continua as measured by force of handgrip. *J. exp. Psychol.*, 1960, 59, 60-67.
- Stevens, S. S. Calculation of the loudness of complex noise. *J. Acoust. Soc. Amer.*, 1956a, 28, 807-832.
- Stevens, S. S. The direct estimation of sensory magnitudes-loudness. *Amer. J. Psychol.*, 1956b, 69, 1-25.
- Stevens, S. S. Cross-modality validation of subjective scales for loudness, vibration and electric shock. *J. exp. Psychol.*, 1959, 57, 201-209.
- Stevens, S. S. Duration, luminance, and the brightness exponent. *Percept. & Psychophys.*, 1966a, 1, 96-100.
- Stevens, S. S. Power group transformations under glare, masking, and recruitment. *J. Acoust. Soc. Amer.*, 1966b, 39, 725-735.
- Stevens, S. S., & Galanter, E. Ratio scales and category scales for a dozen perceptual continua. *J. exp. Psychol.*, 1957, 54, 377-411.
- Stevens, S. S., & Guirao, Miguelina. Subjective scaling of length and area and the matching of length to loudness and brightness. *J. exp. Psychol.*, 1963, 66, 177-186.
- Stevens, S. S., & Harris, Judith R. The scaling of subjective roughness and smoothness. *J. exp. Psychol.*, 1962, 64, 489-494.
- Stewart, M. R., Fagot, R. F., & Eskildsen, P. R. Invariance tests for bisection and fractionation scaling. *Percept. & Psychophys.*, 1967, 2, 323-327.
- Winer, B. J. *Statistical principles in experimental design*. New York: McGraw-Hill, 1962.
- Zwislowski, J., & Hellman, R. On the psychophysical law. *J. Acoust. Soc. Amer.*, 1960, 32, 924.

Notes

1. This research was supported in part by National Science Foundation Grant GB4966, and in part by the Advanced Research Projects Agency of the Department of Defense and was monitored by the Air Force Office of Scientific Research under Contract No. F44620-67-C-0099.
2. Made available by the Statistical Laboratory and Computing Center, University of Oregon.
3. Unfortunately, since the two deviation mean squares are not independent, an F test of the null hypothesis that the two variances are equal is not valid.
4. We are not inclined to agree with their dismissal of random deviations, for it is the characterization of the nature of the random deviations from individual functions which will provide a statistical model for the data and consequent methods for constructing tests and estimating parameters.

(Accepted for publication December 14, 1967.)