

## Using contextual effects to derive psychophysical scales\*

MICHAEL H. BIRNBAUM†

University of California, Los Angeles, Los Angeles, California 90024

This paper presents a functional measurement analysis of Parducci's range-frequency theory. The theory provides a basis for (1) finding context-invariant psychophysical scales, (2) establishing the validity of rating scales, and (3) explaining contextual effects in judgment. In an experimental illustration, Ss judged the magnitude of numerals in nine different distributions. All data were used to obtain the psychophysical function for numerals and also to test the model. The large contextual effects were consistent with the model. Quantitative tests of fit supported the equal-interval assumption of the category scale. The psychophysical function appeared nearly linear but with a significant negative acceleration. The data were shown to be qualitatively inconsistent with generalizations of Helson's theory of adaptation level and Johnson's correlation-regression theory. They supported Parducci's range-frequency theory and illustrated how it could be used to factor out contextual effects from effects of stimulus magnitude. Extensions of the range-frequency approach are also discussed.

Contextual effects are usually considered anathema to direct scaling methods, sources of error and confusion that should be avoided. Magnitude estimation was initially thought to be relatively free of contextual effects, and this supposed invariance was considered a point in its favor (Stevens & Galanter, 1957; Galanter, 1962). However, magnitude estimations, category ratings, and absolute estimations all appear to be subject to such contextual effects (Poulton, 1968; Parducci, 1963; Helson & Kozaki, 1968; Harvey & Campbell, 1963). Luce and Galanter (1963) have concluded: "A full understanding of these effects cannot be expected until we have a sophisticated theory of category judgments. Unfortunately, what is now available is not fully satisfactory. Basically, the problem is to find a response theory which defines a scale of sensation that is invariant under the various experimental manipulations we have just described and does not depend upon an arbitrary, albeit conventional, labeling of the responses."

Such a sophisticated theory of category judgments has been developed. Range-frequency theory (Parducci, 1963, 1965; Parducci & Perrett, 1971) has had considerable success in explaining contextual effects in category ratings. The present paper gives a functional measurement (Anderson, 1970) analysis of the theory to

\*The author would like to thank Allen Parducci, whose research inspired the present work and who provided valuable guidance on this project. Thanks are also due to Norman H. Anderson and Clairice T. Veit for their helpful suggestions on earlier drafts. This research was completed while the author held a National Defense Education Act Title IV graduate fellowship at the University of California, Los Angeles, where computing support was provided by Campus Computing Network. This paper was completed while the author held a National Institute of Mental Health postdoctoral fellowship at the Center for Human Information Processing, University of California, San Diego, which provided assistance in the preparation of the manuscript through NIMH Grant MH-15828.

†Requests for reprints should be addressed to Michael H. Birnbaum, Department of Psychology, Kansas State University, Manhattan, Kansas 66506.

emphasize that it not only explains contextual effects, but can also validate the numerical labeling of the responses and yield context-invariant psychophysical scales.

### Outline of Psychophysical Judgment

Figure 1 provides a basic framework for psychophysical judgment that will facilitate presentation of theories of contextual effects. Psychophysical judgment is interpreted as the composition of two functions: (a) the psychophysical function,

$$\Psi = H(\phi), \quad (1)$$

relating psychological impressions ( $\Psi$ ) to the physical measures of the stimuli ( $\phi$ ); (b) the response function,

$$R = J(\Psi), \quad (2)$$

relating responses ( $R$ ) to the impressions. The function ordinarily observed in a psychophysical experiment is the composition,  $R = J[H(\phi)]$ . As Treisman (1964) has pointed out, different assumptions about  $J$  lead to different conclusions about  $H$ . The present approach is to provide a testable model for  $J$ , rather than assume a particular form for  $J$  as an untested axiom.

Figure 1 also shows  $f(\phi)$ ,  $g(\Psi)$ , and  $p(R)$ , the probability density functions defined on the physical, psychological, and response dimensions. The corresponding cumulative density functions will be denoted,  $F(\phi)$ ,  $G(\Psi)$ , and  $P(R)$ , respectively. The functions  $H$  and  $J$  are assumed to be continuous and strictly monotonic increasing; therefore, their inverse (denoted  $H^{-1}$  and  $J^{-1}$ ) exist. The minimal values on the physical, psychological, and response dimensions are  $\phi_0$ ,  $\Psi_0$ , and  $R_0$ , respectively, and the maximal values are  $\phi_m$ ,  $\Psi_m$ , and  $R_m$ . It is useful to note that  $F(\phi_0) = 0$  and that  $G[H(\phi)] - G[H(\phi_0)] = F[\phi] - F[\phi_0]$ .

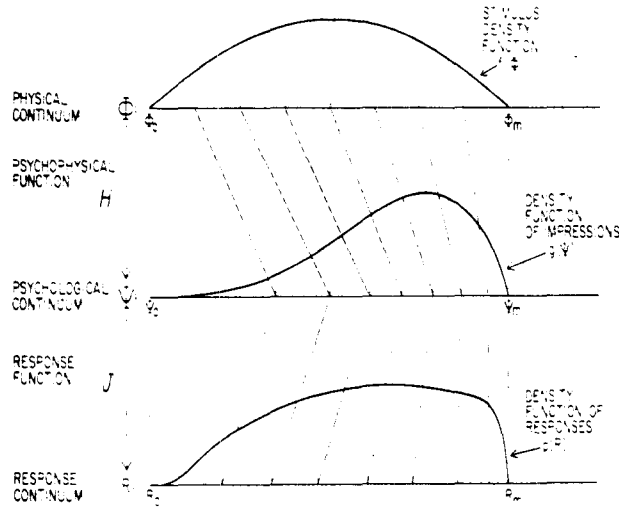


Fig. 1. Framework for discussing contextual effects. In this example, the density function,  $f(\phi)$ , is quadratic, the psychophysical function,  $H(\phi)$ , is a power function, and the judgment function,  $J$ , is assumed to follow range-frequency theory.

**Range-Frequency Theory**

Parducci (1963, 1965, 1973) and Parducci and Perrett (1971) describe psychophysical judgment as a compromise between two principles: (a) the *range* principle asserts that differences in response tend to be proportional to differences in psychological magnitude; (b) the *frequency* principle asserts that differences in response tend to be proportional to differences in stimulus rank.

The *range* principle assumes that differences in response are directly proportional to differences in subjective value and inversely proportional to the range of subjective values. This principle can be written as follows:

$$dR = \left( \frac{b}{\Psi_m - \Psi_0} \right) d\Psi.$$

Subjects tend to locate each stimulus relative to the subjective end values.

The *frequency* principle can be written,  $dR = ag(\Psi)d\Psi$ . Subjects tend to use equal portions of the response continuum with equal frequency.

The actual response is assumed to reflect a weighted compromise of these tendencies. A final empirical principle asserts that differences in response are proportional to the range of responses ( $R_m - R_0$ ) specified by the experimental instructions:

$$dR = (R_m - R_0) \left[ ag(\Psi) + \left( \frac{b}{\Psi_m - \Psi_0} \right) \right] d\Psi, \quad (3)$$

where  $R_m$  and  $R_0$  are the maximal and minimal values of the response and  $a$  and  $b$  are the weights of the

frequency and range principles, respectively.

Integrating Eq. 3 leads to the following statement of the model:

$$R = J(\Psi) = (R_m - R_0) \int_{\Psi_0}^{\Psi} \left[ ag(t) + \left( \frac{b}{\Psi_m - \Psi_0} \right) \right] dt + R_0. \quad (4)$$

The  $H$  function is calibrated to the stimulus range by defining  $\Psi_0 = 0$  and  $\Psi_m = 1$ . The range-frequency approach assumes that  $H$  is invariant when the stimulus endpoints ( $\phi_m$  and  $\phi_0$ ) are held constant.

When the psychophysical function is strictly monotonic and error-free and the stimulus endpoints are fixed, it follows that

$$R = (R_m - R_0) [aF(\phi) + bH(\phi)] + R_0. \quad (5)$$

Equation 5 says that the judgment function is simply a weighted sum of the cumulative density function on the stimulus dimension and the psychophysical function. When the number of categories and the stimulus range are held constant, as in the present experiments, it is a simple matter to vary  $F$  and then solve for  $H$ .

It is useful to modify the notation slightly at this point in order to explain the stimulus scaling and the test of the model. Let  $F_i(\phi)$  be the cumulative density function for context  $i$ ; in practice,  $F_i(\phi)$  is the rank of stimulus  $\phi$  in context  $i$ , divided by the total number of ranks. Let  $R_{\phi i}$  be the judgment of stimulus  $\phi$  in context  $i$ . From Eq. 5, it follows that

$$R_{\phi i} - \alpha F_i(\phi) = \beta H(\phi) + \epsilon, \quad (6)$$

where  $\alpha$  and  $\beta$  are constants and  $\epsilon$  contains response variability. Equation 6 implies that once a best-fit value of  $\alpha$  is found, the differences due to context ( $i$ ) can be eliminated, leaving a context-free psychophysical function. It should be noted, however, that if the model is inappropriate, or if the  $H$  function is not invariant, or if the numerical assignment to the categories is inappropriate, then the model will fail in this test of fit.

The following experiment illustrates the theory and provides a concrete example of how contextual effects may be used to derive psychophysical scales. The stimuli, instructions, and procedure are similar to those of Parducci, Calfee, Marshall, and Davidson (1960). The results for this continuum, although cleaner, are representative of those obtained with other psychophysical continua (Parducci, 1963).

**METHOD**

**Stimuli**

The stimulus context for each  $S$  consisted of a set of from 45 to 47 numerals mimeographed on a single 21 x 35 cm page. The numerals were arranged in a single column at the left side of the

page in order of increasing magnitude. Instructions were printed on the same page. The numerals in each set ranged from 108 to 992, and were spaced according to one of nine density functions corresponding to the nine experimental conditions. For simplicity, the stimulus density functions were chosen as polynomials. Five of the density functions were linear, but with different slopes (-0.022, -0.011, .00, .011, .022). The respective labels for these five conditions vary from -2A through 2A (with A = .011). The 45 numerals used for each of these linear conditions are indicated by the tick marks at the bottom of Fig. 2. Two of the stimulus density functions were approximately quadratic functions, with quadratic coefficients of opposite sign. These two conditions are labeled -B and B, and the 47 particular numerals used are indicated in Fig. 3. The 46 numerals used for each of the remaining two cubic conditions (-C and C) are indicated at the bottom of Fig. 4.

**Subjects**

The 270 Ss were fulfilling a requirement in introductory psychology at the University of California, Los Angeles. There were 30 different Ss in each of the nine conditions, run in groups of from 4 to 10 Ss each. A small number of additional Ss failed to follow instructions, and their data were not tabulated.

**Procedure**

The instructions stated: "This is a study of how people judge the relative magnitude of numbers. On the side of this page, there is a list of numbers ranging in size from *very very small* (108) to *very very large* (992). Your job is to study the entire list and then decide how large or small each number is—in comparison with all the others on the page. Write down (on the space provided beside each number) how large it appears to you. For each number, choose *one* of the following sizes: VVL... Very Very Large, VL... Very Large, L... Large, SLA... Slightly Larger than Average, A... Average, SSA... Slightly Smaller than Average, S... Small, VS... Very Small, VVS... Very Very Small. Be sure to write one of the

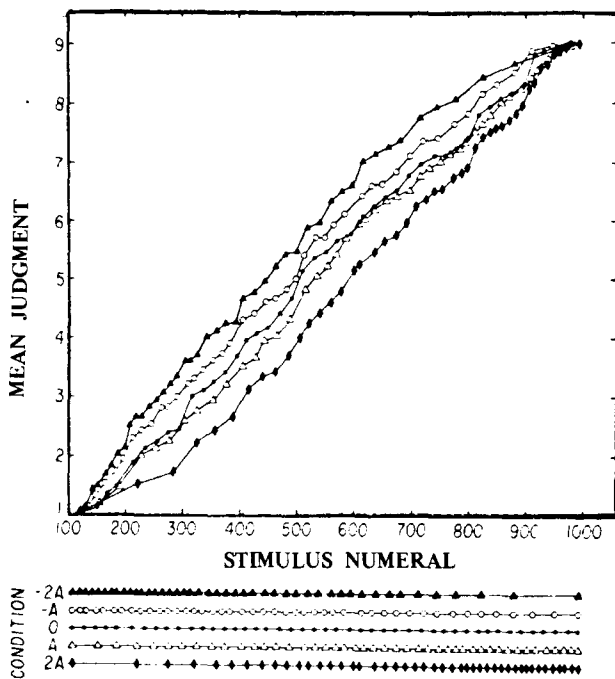


Fig. 2. Mean category judgments for the five linear conditions of stimulus density, -2A to 2A.

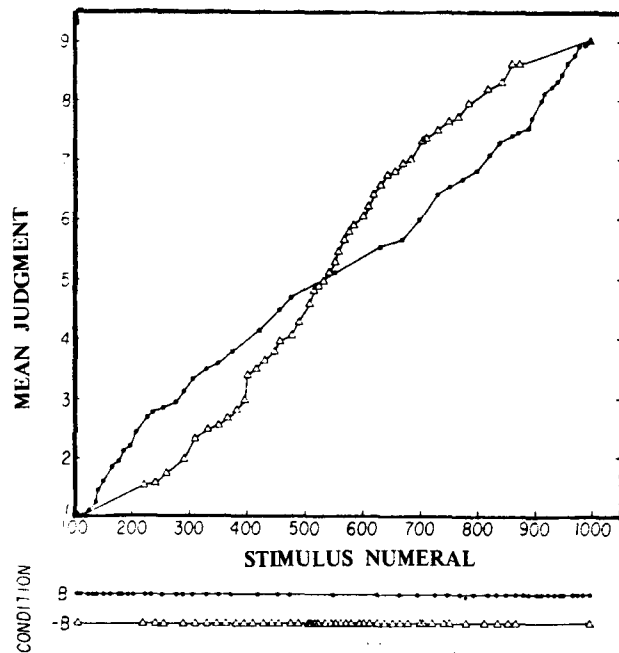


Fig. 3. Mean category judgments for the quadratic conditions of stimulus density, -B and B.

nine abbreviations next to each number. Take your time with these judgments, but do not try to formulate precise rules or to perform numerical calculations."

After recording their judgments of each numeral, Ss turned the page and read instructions which required them to estimate the "typical number" for each category. These instructions specified that Ss were not restricted to the printed numerals but were free to select any numerical values, consistent with their previous judgments, that best represented each of the nine categories.

In summary, the S assigned a category response to each stimulus, and then estimated the "typical stimulus" for each category. Thus, there are two dependent variables, category judgments and "typical number" estimations.

**RESULTS**

The category responses were converted to numerical scores by assigning successive integers from 1 to 9 to the respective categories. The present model provides a basis for testing this numerical assignment and for rescaling, if necessary. The mean judgments are plotted in Figs. 2, 3, and 4 as a function of the stimulus value. Contextual effects appear as differences between the functions. If the judgments were determined by H alone, then all of the curves would coincide. Instead, the curves differ radically in form. The large contextual effects are in the predicted direction, with each judgment function having the approximate form of the cumulative stimulus density function,  $F(\phi)$ .

The "typical number" estimations were averaged and plotted on the abscissa against the integers 1 to 9 on the ordinate for the nine categories. The data for the nine conditions yielded graphs very similar in form to Figs. 2, 3, and 4. The conclusions from the analyses described

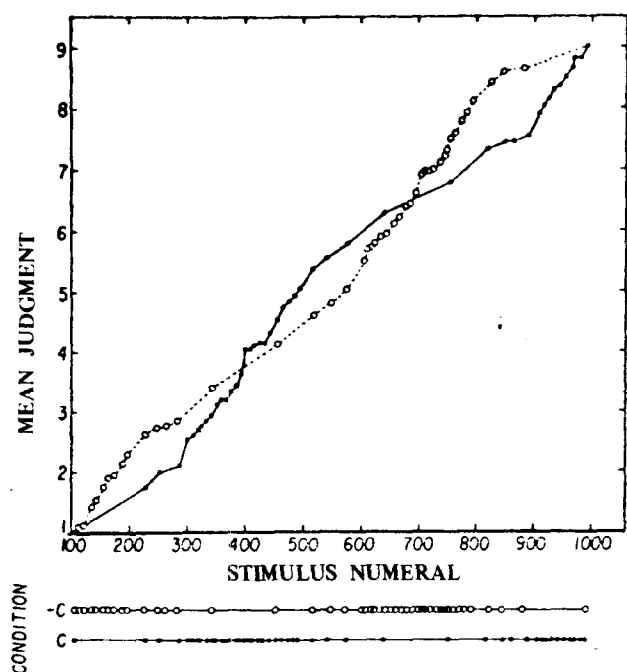


Fig. 4. Mean category judgments for the cubic conditions of stimulus density, -C and C.

below were comparable for both the mean judgment functions and the "typical number" functions.

Assessment of the Range-Frequency Model

According to the theory, all nine functions in Figs. 2-4 may be reduced to a single psychophysical function. The model can be fit to the data by means of multiple regression to find the best-fit constants in the equation,  $R = \alpha F(\phi) + \beta_0 + \beta_1 \phi + \beta_2 \phi^2 + \beta_3 \phi^3$ , in which the H function is approximated as a polynomial. Once the best-fit value of  $\alpha$  is found, Eq. 6 implies that the effects of stimulus distribution can be removed from each of the functions, leaving a context-free psychophysical function.

Figure 5 shows the psychophysical function for numerals derived in this manner from the "typical number" data. On the ordinate is plotted the category ratings corrected for the effect of context—i.e., the residual,  $R - \alpha F_i(\phi_{ki})$ , where the value of R is taken to be k for Category k, and  $\phi_{ki}$  is the mean "typical number" for Category k in Context i. The best-fit value of  $\alpha$  is 3.80. As can be seen in Fig. 5, the residuals from the nine functions closely approximate the same function (dotted curve):

$$\beta H(\phi) = .06 + 6.40 \times 10^{-3} \phi - 8.81 \times 10^{-10} \phi^3. \quad (7)$$

The function appears nearly linear, but the negative acceleration is highly significant. The mean squared error from this function is only .019, which represents less than 3/10 of 1% of the total variance. This excellent fit requires just four parameters: the three constants of

Eq. 7 and  $\alpha$ . Since  $\alpha = a(R_m - R_0)$ ,  $a = .475$ , which is close to the .45 weighting of the frequency principle previously reported (Parducci et al, 1960).

The largest errors in Fig. 5 appear to be for the linear conditions, especially for Condition A. Figure 2 shows the source of the difficulty: the ratings for this condition do not depart as much from the ratings for the uniform distribution as predicted by the model. Analysis of variance was applied to assess the discrepancy for the five linear conditions, the typical-stimulus estimations forming a context by category factorial design. The effects of context and also of the Context by Category interaction were highly significant,  $F(4,145) = 31.72$  and  $F(32,1160) = 8.27$ , respectively, consistent with the model. When the best-fit predictions are subtracted from each S's estimations, analysis of variance detected no significant residual effect: F for all of the pooled sources, with 4 and 1305 df, was only 1.16. Considering the power of the test, the model appears to give a very good account of the data.

Two discrepancies were detected in the corresponding analysis of the mean judgments where the mean squared error was also .019. The first is not really a deviation from the model, but represents a steepening of the H function for numerals in the region of even hundreds. Thus, the judgment functions are steeper in the regions of 300, 400, 500, 600, etc., in Figs. 2, 3, and 4. The second deviation was a tendency for Ss to use Categories "Small" and "Large" with greater relative frequency.

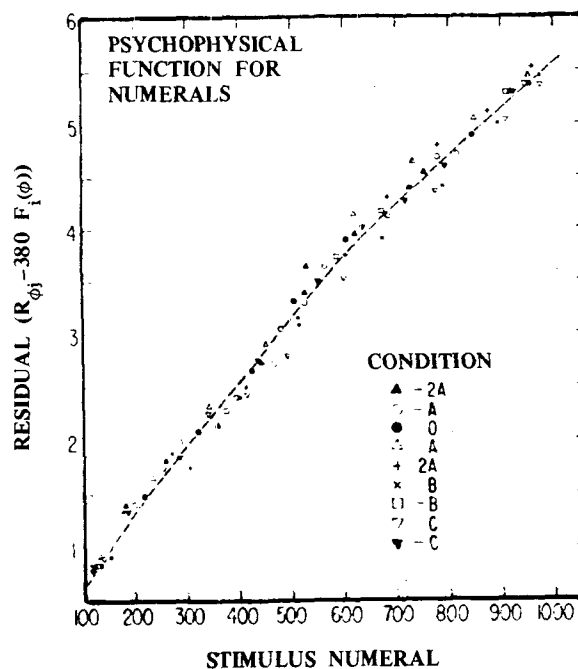


Fig. 5. Context-invariant psychophysical function for numerals. Abscissa values represent the mean "typical number" for each of the nine successive categories; ordinate values are the ratings corrected for the effects of context, i.e., the residual,  $R - \alpha F_i(\phi)$ . The fact that points for different conditions fall on the same curve supports the model, the category scale, and the contextual invariance of the stimulus scale.

This appeared in separate plots of the  $p(R)$  distributions for the different contextual conditions. These sources of error are not serious shortcomings of the model. In a later section, the model will be extended to allow for these effects.

Additional multiple regressions were run to check the possibility that the errors about the fitted psychophysical function in Fig. 5 might be systematically related to higher order powers of  $\phi$  and  $F_i(\phi)$  or their crossproducts. The next few higher powers of  $\phi$  and  $F_i(\phi)$  did not add significantly to the prediction, nor did inclusion of the interaction terms. Thus, it appears that the excellent fit of the simple model for these data is not significantly improved by adding these complications.

## DISCUSSION

The test of fit provides an evaluation of three simultaneous assumptions: that the model, stimulus scale, and response scale are valid. If one of these premises were erroneous, then deviations would have appeared in the test of fit. In this sense, the lack of systematic deviations supports three contentions: (1) It validates the numerical assignment, showing that the response categories are, in fact, psychologically equidistant. (2) It finds a psychophysical function that is invariant with respect to contextual effects and demonstrates this invariance directly. This provides an interval scale of the stimulus. (3) It supports the range-frequency explanation of contextual effects. These three problems will be discussed in turn.

### The Response Scale

Category rating scales have been viewed with suspicion, partly owing to the problem of determining what the appropriate numerical coding of the categories should be. For example, Luce and Galanter (1963) have expressed this skepticism as follows: "To the theorist, however, the whole business [of mean category scales] is a bit hair-raising. To calculate the means of category labels, to plot them against physical measures of the stimuli, and then to discuss the form of the resulting function strikes him as close to meaningless. Because there is nothing about the procedure to prevent one from labeling the categories by any other increasing sequence of numbers, we can by our choice of labels produce any arbitrary monotonic function of the physical stimuli we choose."

The present approach provides a way to determine if the numerical assignment is correct, and related techniques offer a way to transform these initial values to the appropriate values, if necessary. Equation 7 is a special case of an additive model; therefore, any nonlinear transformation of the responses would alter the fit of the model. For example, letting  $R = k^2$  for the  $k^{\text{th}}$  category in the present case led to a failure of the

model for the present data. The reason for this can be seen from the appearance of an interaction term when both sides of Eq. 5 are squared.

If the a priori numerical coding of the responses does not fit the model, techniques such as those described by Bogartz and Wackwitz (1971) may be used to find the rescaling which makes the model fit. No rescaling was necessary for the present data, indicating that for at least the present conditions, verbal categories provide an equal interval scale.

### The Stimulus Scale

Obviously, category rating scales do not represent the psychophysical function directly, as illustrated by the gross differences in Figs. 2, 3, and 4. The category ratings may be an equal interval scale of the responses, but need not correspond to the psychophysical function, defined as  $H$  in the model. It is the latter, stimulus scale, that is invariant with respect to the stimulus distribution.

The meaning of a psychophysical scale, aside from its role in fitting a single model, is contained in its ability to predict to a variety of other situations. Hence, the psychophysical scales derived from the fit of two or more different models in different situations using the same stimuli should agree (Birnbaum, 1972; Birnbaum & Veit, 1973).

A recent experiment by Rose and Birnbaum<sup>1</sup> provides convergent support. In that experiment, Ss divided a line segment to represent the difference in magnitude between pairs of numerals. The scale values derived in accordance with a subtractive model were a nearly perfect linear function of the scale values derived from the present data (Eq. 7). Other investigators, using somewhat different approaches, have also concluded that the psychophysical function for numerals is negatively accelerated (see Rule & Curtis, 1973).

### The Model

#### *Range-Frequency Theory*

The present data add further support to Parducci's range-frequency theory, which has had considerable quantitative success in predicting ratings of unidimensional stimuli presented in varying stimulus distributions (Parducci et al, 1960; Parducci, 1963, 1965; Parducci & Perrett, 1971). The model gives a very good account of the radically different functions shown in Figs. 2, 3, and 4, using just four parameters. Figure 5 shows that in spite of the differences in the judgment function for different contexts, the data from all of the distributions lie on a common curve. Hence, the psychophysical function implied by the model is invariant with respect to context.

Much of Parducci's research has been directed at providing critical tests between range-frequency theory

and alternative approaches. The present data provide illustrations of these tests. It is instructive to show how alternative models fail to account for qualitative features of the data.

#### *Adaptation-Level Theory*

Adaptation-level theory (Helson, 1964) interprets psychophysical judgment as a process of comparison in which each stimulus is compared with the "average" stimulus—the adaptation level (AL). Every influence on judgment is considered a stimulus, and all stimuli pool to form a single quantity (AL), which serves as the point of reference for judgment. It will be shown that even the most general statement of this theory cannot account for the crossovers in Figs. 3 and 4.

Adaptation-level theory can be generalized in the framework of Fig. 1 as follows:

$$R_{\phi i} = U(\Psi_{\phi} - \tilde{\Psi}_i), \quad (8)$$

where  $R_{\phi i}$  is the response to stimulus  $\phi$  in context  $i$ ,  $U$  is a continuous monotonic, strictly increasing function,  $\Psi_{\phi}$  is the psychological value of stimulus  $\phi$ , and  $\tilde{\Psi}_i$  is the psychological value of the AL for Context  $i$ . In the present analysis, it is unnecessary to make any of the usual adaptation-level theory assumptions that  $\Psi_{\phi} = \log(\phi)$ , that  $\tilde{\Psi}_i$  is a weighted average of the psychological values of the stimuli, background, and residual influences, or that  $U$  is linear.

Equation 8 can be rewritten  $U^{-1}(R_{\phi i}) = \Psi_{\phi} - \tilde{\Psi}_i$ . Therefore, it implies that it should be possible to rescale the ordinates of Figs. 2, 3, and 4 so that the curves for the different context conditions will be parallel. Clearly, no possible monotone transformation can eliminate the Context by Stimulus interactions. Indeed, the curves in Figs. 3 and 4 show the crossover interactions (also shown by Parducci, 1963, and Parducci & Perrett, 1971, for traditional perceptual continua) which are incompatible with even this most general form of adaptation-level theory.

The implication that crossovers are inconsistent with adaptation-level theory, i.e., that

$$R_{\phi k} > R_{\phi i} \text{ if and only if } R_{\phi' k} > R_{\phi' i},$$

can be easily proved as follows: The left-hand side is equivalent to  $U(\Psi_{\phi} - \tilde{\Psi}_k) > U(\Psi_{\phi} - \tilde{\Psi}_i)$ , which is equivalent to  $\Psi_{\phi} - \tilde{\Psi}_k > \Psi_{\phi} - \tilde{\Psi}_i$ ; thus,  $-\tilde{\Psi}_k > -\tilde{\Psi}_i$ ; addition of  $\Psi_{\phi'}$  to both sides of the equation yields  $\Psi_{\phi'} - \tilde{\Psi}_k > \Psi_{\phi'} - \tilde{\Psi}_i$ , which is equivalent to  $R_{\phi' k} > R_{\phi' i}$ . It should be clear that a ratio formulation of adaptation-level theory, with the assumption of positive scales, makes the same incorrect prediction.

#### *Correlation-Regression Theory*

Johnson and Mullaly (1969) have proposed a model

that views the stimulus-response relationship in psychophysical judgment as an example of statistical regression. The correlation-regression model can be reformulated as follows:

$$R_{\phi i} = \frac{r_{\Psi R} \sigma_R}{\sigma_{\Psi_i}} (\Psi_{\phi} - \tilde{\Psi}_i) + \bar{R}_i, \quad (9)$$

where  $\bar{R}_i$  is the middle response,  $\sigma_R$  is the standard deviation of the response scale,  $\Psi_{\phi}$  is the psychological value of stimulus  $\phi$ ,  $\tilde{\Psi}_i$  is the mean of the psychological values of the stimuli in context  $i$ ,  $\sigma_{\Psi_i}$  is the standard deviation of the psychological values of the stimuli in context  $i$ , and  $r_{\Psi R}$  is the correlation between the psychological values and the responses (an empirical measure of the unreliability of judgment). This approach correctly describes the fact that the slope of the judgment function varies inversely with the spread of the stimuli and directly with the reliability of judgment.

The regression model assumes that judgments are linear functions of the psychological values of the stimuli, with different slopes and intercepts for the different conditions—the slope being inversely proportional to the spread of the psychological values of the stimuli and the intercept inversely related to their mean. It should therefore be possible to find a monotonic transformation of the abscissas of Figs. 2, 3, and 4 that would render the judgments to be linear functions. Again, it should be clear that no transformation will work. Two straight lines can intersect at most at one point; therefore, the intertwining functions of Fig. 4 cannot be explained by their theory. Such intertwining functions have been employed by Parducci (1963, 1965) and Parducci and Perrett (1971) to disprove this entire class of models that dictate linear relationships between rescaled values for the stimuli and the responses.

### **Extensions of the Range-Frequency Model**

#### *The Frequency Principle*

It is the frequency principle of range-frequency theory that allows the theory to account for the radically different judgment functions. However, this principle can be generalized to allow for the case in which Ss may have a tendency to use their categories with some fixed, but not necessarily uniform, distribution. Suppose the numerals in the present study represented examination scores which will determine a student's grade in a class—the rating categories would be grades, such as A, B, C, D, and F. Under these circumstances, the Ss would probably exhibit a tendency to assign more Cs than any other category, and to give more As and Bs than Ds and Fs. Indeed, university departments have been known to issue guidelines calling for skewed distributions of grades. However, differences

in grade will also depend upon differences in examination scores.

In terms of the model, equal frequency implies that only linear terms in  $F_i(\phi)$  are necessary to predict the rating; nonequal frequency would involve some function of  $F_i(\phi)$ . Thus, the model can be generalized as follows:

$$R_{\phi_i} = \beta H(\phi) + Q[F(\phi)], \quad (10)$$

where  $Q$  is some monotonic function that might in practice be approximated as a polynomial. It can be shown that the nonlinear terms in  $Q$  give an approximation of the frequencies with which the subjects "tend" to use different categories. In particular, the derivative of the  $Q$ -inverse function describes the frequency distribution of category usage.

Equation 10 was applied to the mean ratings of the sizes of squares reported by Parducci and Perrett (1971): the coefficient of  $F^2(\phi)$  was significantly positive, indicating a tendency for  $S_s$  to use their lower categories with greater frequency. The improvement over the simple analysis of Parducci and Perrett (which assumes an equal-frequency tendency and equal weighting of range and frequency principles) was not overwhelming, but it was regular.

It is also noteworthy that the issue of nonuniform frequency usage is theoretically separable from the response rescaling issue. If the additive model of Eq. 10 fits the data, then the numerical coding of the responses is appropriate, whatever the nature of  $Q$ . Equation 10 can be tested graphically by plotting the responses as a function of  $F_i(\phi)$ , with a separate curve for each stimulus presented in sets with the same endpoints. The curves should be parallel, but need not be linear.

For the grading example, it may be the case that letter grades are not psychologically equidistant *and* subjects have a nonuniform response tendency. In this case, the assumption of Eq. 10 and the use of the Bogartz and Wackwitz (1971) technique would lead to a scaling of the psychological values of the grades.

#### Range Principle

If the psychophysical function is to have greater generality, the psychophysical function derived in one stimulus range should be linearly related to the psychophysical function derived in another range. For judgments of the same stimuli having the same ranks in different contexts with varying endpoints, the model can be written:

$$R_{\phi_i} = \alpha F(\phi) + \beta \left[ \frac{\Psi - \Psi_{0_i}}{\Psi_{m_i} - \Psi_{0_i}} \right], \quad (11)$$

where  $\Psi_{m_i}$  and  $\Psi_{0_i}$  are the subjective values of the greatest and smallest stimulus in context  $i$ . This shows that judgments of the same stimuli having the same rank

in different contexts should be linearly related. This prediction can be investigated by testing the residual from bilinearity (see Anderson, 1970). Data of Parducci and Perrett (1971) appear to follow this prediction roughly, although the research was not designed to permit an ideal test.

#### Concluding Comment

The context-invariant psychophysical function should not draw attention away from the contextually dependent ratings. Neither candidate has a greater claim to be the "true" psychological scale. In order to predict ratings in any given context, it is necessary to know both the psychophysical function and the contextual theory. Historically, "context effects" were considered undesirable in scaling experiments and were labeled "biases" and "noise." However, the lawfulness of these effects demonstrates that they are a substantive psychological issue of equal or greater importance than the psychophysical scales that can be derived from them.

Contrary to the view that contextual effects may represent a challenge to the basic premise of scaling, this paper contends that theories of contextual effects can provide a practical basis for scaling. A popular view in psychophysics is that contextual effects are a "nuisance" that should be "averaged out" or "avoided" through standardized experimental procedures. The present view contends that *systematic manipulation of the context*, together with *appropriate theorization*, can lead to context-invariant psychophysical scales. Such manipulation permits a test of the theory while demonstrating the invariance directly.

#### REFERENCES

- Anderson, N. H. Functional measurement and psychophysical measurement. *Psychological Review*, 1970, 77, 153-170.
- Birnbaum, M. H. The nonadditivity of impressions. Unpublished PhD dissertation, University of California, Los Angeles, 1972.
- Birnbaum, M. H., & Veit, C. T. Scale convergence as a criterion for rescaling: Information integration with difference, ratio, and averaging tasks. *Perception & Psychophysics*, 1974, 15, 7-15.
- Bogartz, R. S., & Wackwitz, J. H. Polynomial response scaling and functional measurement. *Journal of Mathematical Psychology*, 1971, 8, 418-443.
- Galanter, E. Contemporary psychophysics. In R. Brown et al (Eds.), *New directions in psychology I*. New York: Holt, Rinehart, & Winston, 1962.
- Garner, W. R. Context effects and the validity of loudness scales. *Journal of Experimental Psychology*, 1954, 48, 218-224.
- Harvey, O. J., & Campbell, D. T. Judgment of weight as affected by adaptation range, adaptation duration, magnitude of unlabeled anchor, and judgment language. *Journal of Experimental Psychology*, 1963, 65, 12-21.
- Helson, H. *Adaptation-level theory*. New York: Harper & Row, 1964.
- Helson, H., & Kosaki, A. Anchor effects using numerical estimates of simple dot patterns. *Perception & Psychophysics*, 1968, 4, 163-164.

- Johnson, D. M., & Mullally, C. R. Correlation-and-regression model for category judgments. *Psychological Review*, 1969, 76, 205-215.
- Luce, R. D., & Galanter, E. Psychophysical scaling. In R. D. Luce, R. R. Bush, and E. Galanter (Eds.), *Handbook of mathematical psychology*. Vol. 1. New York: Wiley, 1963.
- Parducci, A. Range-frequency compromise in judgment. *Psychological Monographs*, 1963, 77(2, Whole No. 565).
- Parducci, A. Category judgment: A range-frequency model. *Psychological Review*, 1965, 72, 407-418.
- Parducci, A. Contextual effects: A range-frequency analysis. In E. C. Carterette and M. P. Friedman (Eds.), *Handbook of perception*. Vol. II. New York: Academic Press, 1973, in press.
- Parducci, A., Calfee, R. C., Marshall, L. M., & Davidson, L. P. Context effects in judgment: Adaptation-level as function of mean, midpoint, and median of the stimuli. *Journal of Experimental Psychology*, 1960, 60, 65-77.
- Parducci, A., & Perrett, L. F. Category rating scales: Effects of spacing and frequency of stimulus values. *Journal of Experimental Psychology Monograph*, 1971, 89, 427-452.
- Poulton, E. C. The new psychophysics: Six models for magnitude estimation. *Psychological Bulletin*, 1968, 69, 1-19.
- Rule, S. J., & Curtis, D. W. Conjoint scaling of subjective number and weight. *Journal of Experimental Psychology*, 1973, 97, 305-309.
- Stevens, S. S. Issues in psychophysical measurement. *Psychological Review*, 1971, 78, 426-450.
- Stevens, S. S., & Galanter, E. Ratio scales and category scales for a dozen perceptual continua. *Journal of Experimental Psychology*, 1957, 54, 377-411.
- Treisman, M. Sensory scaling and the psychophysical law. *Quarterly Journal of Experimental Psychology*, 1964, 16, 11-22.

## NOTE

1. Rose, B. J., & Birnbaum, M. H. Cross-modality integration matching with difference and ratio tasks. Submitted for publication.

(Received for publication March 5, 1973;  
accepted August 9, 1973.)