

The Category Effect With Rating Scales: Number of Categories, Number of Stimuli, and Method of Presentation

Allen Parducci and Douglas H. Wedell
University of California, Los Angeles

Squares receive higher category ratings when the smaller sizes are presented more frequently than the larger sizes. This shift in the rating scale is greater when there are either fewer categories (the Category Effect) or more stimuli. Similar shifts were obtained whether the stimuli were presented successively for judgment or simultaneously. The Category Effect also occurred when subjects were not told how many categories to use until after the contextual stimuli had been presented. A simple range-frequency model describes most of the shifts in scale by variations in a single weighting parameter. However, these shifts are predicted by an elaborated model in which the number of representations of any stimulus in working memory is limited by a principle of consistent assignment of each stimulus to a single category. This elaborated model correctly predicts the disappearance of the Category Effect when contexts are manipulated by varying the spacing of stimulus values rather than by varying their relative frequencies.

Category ratings are probably the most frequently used of dependent variables in psychological research, perhaps because they reflect the way people ordinarily express value judgments. In opinion surveys, in personality inventories, and also in a wide variety of laboratory experiments, subjects make ratings using a prescribed set of categories. How many categories? It is commonly believed that the specific number of categories is unimportant, for example, that the scale of judgment for three categories is simply a linear transformation of the scale for nine categories. This assumption of the equivalence of scales is sometimes wrong.

Compare the two panels of Figure 1, which plot the effects of a contextual manipulation on ratings of the same five squares, using either three or nine categories. Both panels show typical contextual effects: Ratings of the same stimuli are higher when the distribution of contextual frequencies is positively skewed (i.e., when the smaller squares occur with greater frequency). However, the effects of this contextual skewing are more than twice as great when subjects are restricted to just *small*, *medium*, and *large* than when they are permitted to use nine different categories from *very, very small* to *very, very large*. This difference in the magnitude of the contextual effect suggests that the number of categories may be a crucial consideration in research using category ratings as a dependent variable. For example, in food research where tasters rate the sweetness of different soft drinks, the same drink might be reported as below *medium sweetness* using one set of categories but above *medium sweetness* using a set with a different number of categories.

How is sensitivity to the skewing of the distribution of contextual stimuli affected by the number of categories and the num-

ber of stimuli? Previous research (Parducci, 1982) supported two empirical generalizations: (a) The effects of skewing diminish progressively from two to nine categories and are even smaller for 20-point and 100-point numerical scales, and (b) the effects of skewing increase with more stimulus levels. These two generalizations, here called the Category Effect and the Stimulus Effect, can be described by changes in the single weighting parameter of range-frequency theory.

Range-Frequency Theory

Range-frequency theory describes how category ratings are determined by the frequency distribution of contextual stimuli. The context for judgment is conceived as a set of stimuli whose spacing and relative frequency of presentation determine how each of its component stimuli is judged. The theory was first developed to account for ratings of the subjective magnitudes of numerals (Parducci, Calfee, Marshall, & Davidson, 1960), where the context for judgment was assumed to consist of the particular set of numerals presented simultaneously on the same page. With psychophysical stimuli presented successively in random order, the different presentations of the same stimulus were assumed to elicit different discriminational processes, as inferred by Thurstone (1927) and the theory of signal detection (Green & Swets, 1966). Although this approach has certain advantages in accounting for the variability of judgment (Parducci, 1965), subsequent research (Parducci & Perrett, 1971) demonstrated that the effects of various manipulations of the distribution of contextual stimuli were fit better by the present model, which ignores the variability of stimulus representation. Because variations in range have a large effect on the variability of judgment, it can be assumed that the discriminational dispersion for each stimulus is relatively small (Braid & Durlach, 1972; Parducci & Perrett, 1971).

The essential idea of range-frequency theory (elaborated in Parducci, 1983) is that the judgment of any particular stimulus represents a compromise between two principles of judgment: (a) The respective categories are assigned to successive, objec-

We are grateful to Michael H. Birnbaum for his constructive criticism of a draft of this article.

Correspondence concerning this article should be addressed to Allen Parducci, Department of Psychology, Franz Hall, University of California, 405 Hilgard Avenue, Los Angeles, California 90024.

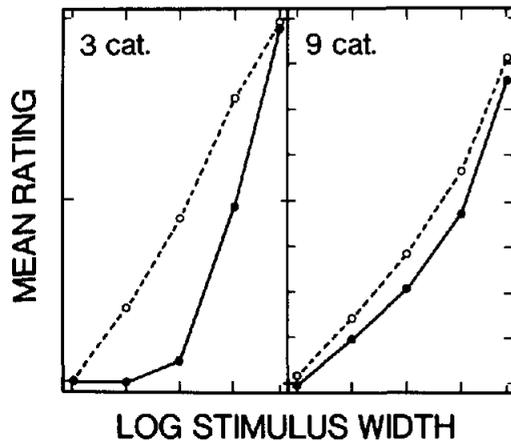


Figure 1. Category Effect: greater effect of contextual skewing for three categories (cat.) than for nine categories. (Stimuli presented with frequencies of 10-7-4-2-2 for positive set [ratings shown by open points], 2-2-4-7-10 for negative set [solid points].)

tively equal subranges of the contextual stimuli (the range principle), and (b) the same number of contextual stimuli are assigned to each of the available categories (the frequency principle). The category rating, C_{ic} , of Stimulus i in Context c is simply a weighted average of these two principles:

$$C_{ic} = wR_{ic} + (1 - w)F_{ic}, \quad (1)$$

where R_{ic} is the *range value* of Stimulus i in Context c (what its rating would have been if categories actually divided the contextual range into subjectively equal subranges); F_{ic} is the *frequency value* of the same stimulus (what its rating would have been if an equal number of contextual stimuli were assigned to each category while maintaining an ordinal scale); and w is the weighting parameter describing the compromise between the two principles of judgment.

In the earliest formulation of range-frequency theory (Parducci et al., 1960), the actual values of the numerals printed on the page were used to determine the range values. For example, the numeral 550 was assumed to be at the midpoint of the subjective range of a contextual set of numerals that extended from 100 to 1,000. In applications to psychophysical judgments, range values were inferred from the data using Equation 1. The assumption that a particular stimulus would have the same range value when presented in different contextual sets with the same two end-stimuli permits powerful tests of the theory (e.g., Birnbaum, 1974; Parducci & Perrett, 1971). Such tests generally assume that *all* stimulus presentations remain in the effective context for judgment so that frequency values can be calculated directly from the contextual distribution actually presented.

The weighting parameter, w , reflects the relative influence of range and frequency principles in judgment. Although it is inferred empirically from overt category ratings, its value has fallen close to .5 in a wide variety of empirical studies using different stimulus dimensions (e.g., Parducci et al., 1960; Parducci & Perrett, 1971; Risky, Parducci, & Beauchamp, 1979). Thus, it is surprising that, as in Figure 1, the effects of skewing the contextual distributions vary so dramatically with different

numbers of categories (the Category Effect), for this suggests dramatic changes in w .

Interpreting Category and Stimulus Effects

Change in weighting. An important question to be answered by the present research is whether Category and Stimulus Effects can be adequately described by variation in w . Insofar as they can, shifts in w might be attributed to increased difficulty in calculating frequency values when presentations of the same stimulus must be divided among a greater number of categories. Such increased difficulty would seem to occur when there are more categories to distribute or fewer stimuli among which to distribute the same number of categories. Because the degree of this kind of difficulty is not easily quantified, the different values of w would have to be determined empirically.

Retrieval of stimulus frequencies. Past methods for calculating frequency values have assumed that all stimulus presentations enter the effective context for judgment and remain effective throughout the experimental session. Although this assumption has resulted in good fits of the range-frequency model to data, it implies an astonishingly accurate retrieval of past presentations. Suppose, instead, that increasing the number of categories limits the number of repetitions that can be retrieved for any given stimulus. The remembered distribution would then become more uniform, with a consequent reduction in the effects of skewing (the Category Effect). The stimulus effect might then be a simple consequence of the decrease in repetitions per stimulus that occur when the number of different stimuli is increased. Wedell (1984) has shown how such a memory-retrieval model can account quantitatively for both Category and Stimulus Effects.

Consistent identification. It is also possible that both Category and Stimulus Effects reflect a principle of consistency: The same category should be assigned to all repetitions of the same stimulus. A functional interpretation of the frequency principle emphasizes its efficiency in *identifying* the stimuli. When the number of distinct stimuli is large relative to the number of categories, assigning an equal number of stimuli to each category maximizes the transmission of information—in the Shannon-Weaver sense of information theory (Attneave, 1959; Garner, 1962). However, when there are only a few different stimuli relative to the number of categories, assigning equal numbers of stimuli to each category (the frequency principle) results in the same stimulus being assigned to different categories. This type of inconsistent identification reduces the transmission of information. As will be evident in Experiment 4, subjects do not violate the principle of consistency when it would be clear that they were doing so—as in certain conditions of simultaneous presentation.

A retrieval-consistency model. How might the consistency principle limit the number of repetitions retrieved for each stimulus? Subjects might search memory for past presentations of any particular stimulus only up to the limit implied by the frequency principle. For example, if nine different stimulus values were being rated using just three categories, the frequency principle would limit the application of each category to one third of all presentations. Retrieving more than one third would bring the frequency principle into conflict with the principle of consistency. However, if nine categories were available for rating

Table 1
Overview of Experimental Conditions

Experiment	Number of		Method of	
	Categories	Stimuli	Presentation	Skewing
1	2-100	5, 9, 24	Successive	Frequency
2	3, 9	5	Successive	Frequency ^a
3	3, 5, 100	5, 9	Successive	Spacing
4A	3, 5, 9	5, 9	Simultaneous	Frequency
4B	3, 9	5, 9 ^b	Simultaneous	Frequency
4C	3, 5, 9, 100	11, 8, 44 ^c	Simultaneous	Mixed

^a Skewed (incidental exposure) followed by uniform distribution. ^bSquares, numerals, or lines in five or nine groups. ^cFor dots, squares, and numerals, respectively.

the same nine stimuli, the conflict would occur after retrieving any presentations of a stimulus that would bring its count above one ninth of all contextual presentations. This type of limitation on the number of stimulus repetitions included in the effective context for judgment would reduce the effects of skewing with either an increase in the number of categories or a decrease in the number of stimuli (given that fewer distinct stimuli entail more repetitions per stimulus).

The consistency principle can thus be incorporated into an elaborated range-frequency model that quantitatively predicts both Category and Stimulus Effects. Drawing upon an earlier study of the effects of transfer from one skewed distribution to another (Wedell, 1984), we assume that retrieval of past frequencies is limited to a search set consisting of only the 12 most recent presentations. Because the consistency principle precludes assignment of more than one category to each stimulus, past presentations of a stimulus are not counted beyond the frequency of use of each category as dictated by the frequency principle. These premises may be summarized as follows:

P1: Search set = stimuli presented on last 12 trials.

P2: Depth of search = search set ÷ number of categories. (Thus the assumed depth of search of two categories is simply $12 \div 2 = 6$; for three categories, it is $12 \div 3 = 4$; etc.)

P3: If presentation frequency \leq depth of search, effective stimulus frequency = presentation frequency.

P4: If presentation frequency $>$ depth of search, effective stimulus frequency = depth of search.

These premises allow the range-frequency model (Equation 1) to account for both the category and stimulus effects using a single value of w .

Overview of the Research

Table 1 provides an overview of the different experiments, all of which compare category ratings of sets of stimuli with different skewing (positive and negative) under different conditions. Experiment 1 demonstrates category and stimulus effects under a variety of conditions. The changes in rating scales are described quantitatively, using either the simple range-frequency model (for which w must be estimated for each condition) or the elaborated retrieval-consistency model (for which shifts in frequency values can be calculated without reference to the data). Experiment 1 also rules out a number of alternative interpretations of the data. Experiments 2, 3, and 4 explore lim-

iting conditions upon Category and Stimulus Effects implied by memory and consistency interpretations. Experiment 2 attempts to locate the Category Effect at either the encoding or retrieval stage of memory, testing whether memory for past responses is crucial. Experiment 3 determines whether Category and Stimulus Effects disappear when it is the spacing of stimulus values (rather than their distribution of frequencies) that is skewed, an implication of the retrieval-consistency interpretation. Experiments 4A, 4B, and 4C test whether the interpretations used to account for the data for successive presentations also apply when the stimuli are presented simultaneously.

Experiment 1: Number of Categories and Stimuli

Method

Design

The major objective was to determine how the effects of skewing interact with the number of categories, that is, the nature of the Category Effect. There were three major independent variables: (a) skewing of stimulus frequencies, (b) number of categories, and (c) number of stimuli. For each positively skewed condition, there was a corresponding negatively skewed condition that used the same number of categories and was also the same in all other respects (except that different subjects were used). Each subject was run under only one condition because pilot work revealed significant transfer when the same subjects were shifted from one condition to another. Consequently, the only within-subjects factor is the different stimulus values from the same set.

Instructions

Each experimental session began with instructions for judgment. Subjects were asked to read the printed instructions to themselves while the experimenter read them aloud. These described the experiment as research on how people judge the sizes of the different squares to be projected one at a time on the screen before them. Their task, working as individuals, was to judge each square in accordance with how large or small it appeared, in comparison with the other squares. For each presentation, they were to record one of the categories of judgment printed at the top of their response sheet. A preview would be presented first to familiarize them with the different sizes, but they were to record judgments right from the beginning in order to get into the rhythm of the presentations. The instructions stipulated that they were to record only the initials of the categories or (for different subjects) a number from the numerical rating scale. For the open conditions, subjects were

Table 2
Stimulus Frequency Distributions Used in Experiment 1

Set	Stimuli																							
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
Positive																								
5 stim	20						14						8					4						4
9 stim	10			10			12		4				4			4			2			2		2
24 stim	6	4	6	4	4	4	4	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2
Negative																								
5 stim	4						4						8						14					20
9 stim	2		2			2			8			8			8			12			10			10
24 stim	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	4	4	4	4	6	4	6

Note. Stim = stimuli.

free to use any categories that seemed natural to describe how large or small the different squares appeared.

Categories. Prescribed sets of verbal categories were provided for subjects in different experimental conditions. The major differences were with respect to the number of categories: two (small and large), three (small, medium, and large), four (small, slightly smaller than average, slightly larger than average, large), five (very small, small, average, large, very large), seven (very small, small, slightly smaller than average, average, slightly larger than average, large, very large), and nine (very very small, very small, small, slightly smaller than average, average, slightly larger than average, large, very large, very very large). In the open conditions, subjects were not told what labels to use or even how many. There were also two additional conditions that used numerical rating scales, 0-20 (with 0 for the smallest size, 20 for the largest) and 0-100 (with 0 for the smallest, 100 for the largest).

Stimuli

All squares were members of a master set of 24 sizes, constituting a geometric series whose successive widths increased by a factor of 1.16. Varying in width from 1.3 cm (Square 1) to 39.5 cm (Square 24), these solid black squares were projected against a 75 × 100 cm white background placed from 1.5 m to 3 m in front of the seated subjects. The projector switched slides every 5 s, with a switching interval of 0.5 s.

Table 2 shows frequency distributions for the 5-, 9-, and 24-stimulus, positive and negative sets. Within each set, stimuli were spaced at approximately equal log-physical widths. The frequencies were selected to make the plots of percentile ranks against log stimulus width as independent of the number of stimuli as possible. Although this independence is not crucial when skewing effects are measured by w from Equation 1, it is for the statistical tests of the Stimulus Effect. The plots of the percentile ranks actually showed slightly greater skewing for 5 than for 9 or 24 stimuli (the area within the positive and negative functions was 11.1% and 10.6% greater, respectively). These differences are due to the greater pinching together of the functions with more stimuli and would work against demonstrating the statistical significance of the Stimulus Effect.

Procedure

The method of single stimuli was employed; subjects were presented a random sequence of squares and made category ratings of their successive sizes. There were 10 preview presentations for the 5-stimulus conditions, 12 for the 9- and 24-stimulus conditions. These were selected to be representative of the particular skewing used and included both end-stimuli. After the preview, the experimenter recycled the projector tray and announced that subsequent responses would be tabulated as the real data of the experiment. Presentations were randomized in two

blocks of 25 for the 5- and 9-stimulus conditions, and in two blocks of 50 for the 24-stimulus conditions (with the restriction that each block resemble as closely as possible the total distribution). Two different random sequences were used for each condition so that different groups of subjects received different sequences.

Subjects

All subjects were students enrolled in introductory psychology at the University of California, Los Angeles, receiving credit by serving in the experiment. About 20 different subjects served in each condition, 10 for each of its two sequences. Some conditions were replicated, with an additional 20 subjects per replication.

Dependent Variable

Plotting the data. The basic dependent variable is the category rating assigned by an experimental subject to the presentation of a particular square. In tabulating the data, the experimenter first transformed the letter abbreviation for each verbal category into its rank in the set of prescribed categories (e.g., the rating *S* for *small* was tabulated as 1 for two-category conditions, as 3 for nine-category conditions). To permit comparisons between conditions with different numbers of categories, these ranks were then linearly transformed to a zero-to-one scale:

$$J_{ic} = (C_{ic} - 1)/b, \quad (2)$$

where J_{ic} is the judgment of Stimulus i in Context c normalized to a scale from 0 to 1; C_{ic} is the rank (in the prescribed set of categories) corresponding to the overt category rating of that stimulus; and b is the range of these ranks for the particular number of categories (e.g., 8 for a 9-category condition). This transformation sets the two prescribed end-categories at 0 and 1 for all conditions.

Statistical analyses. Each subject's mean judgment of each stimulus (as defined by Equation 3) was used to calculate the area beneath his or her judgment function, equivalent to linearly interpolating a judgment for each unjudged stimulus and taking the mean of the judgments of all 24 stimuli. Because this mean represents the area beneath the judgment function, differences between the means for positive versus negative sets measure the overall effects of skewing in terms of the area enclosed by their respective judgment functions. This was the dependent variable for all statistical analyses.

Results and Discussion

Demonstration of Category and Stimulus Effects

Five-stimulus conditions. Figure 2 illustrates the interaction between the effects of skewing and number of categories. The

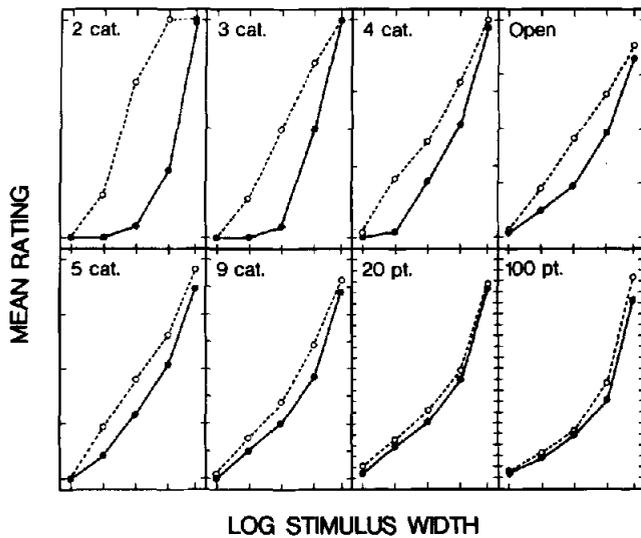


Figure 2. Category Effect: Effects of skewing vary inversely with number of categories—for five-stimulus sets. (Ratings of positive sets shown by open points, negative sets by solid points. Cat. = category; pt. = point.)

Category Effect is indicated by the diminishing differences between scales for positive and negative sets across successive panels. The panels for the five-category and 100-point scales add one complete replication to the data from Parducci (1982), and panels for the three- and nine-category scales add data from two new replications.

The standard error for individual points in Figure 2 averages about .10 of a category step for the five-category conditions (with slightly greater vertical displacement for fewer categories but less for more categories); thus the decline in the effects of skewing with more categories is highly significant. This significant interaction effect was confirmed by an overall 2×8 (Skewing \times Categories) analysis of variance (ANOVA) following the design of Figure 2, $F(7, 640) = 32.52, p < .0001$.

Nine- and 24-stimulus conditions. The lower panels of Figure 3 illustrate the Category Effect for conditions in which nine rather than just five stimuli are presented (the corresponding five-stimulus scales from Figure 2 are presented in the upper panels for comparison). Although the effects of skewing are manifestly larger with nine stimuli (the Stimulus Effect), the diminishing effects of skewing with increase in categories (the Category Effect) also occurs with nine stimuli. These generalizations are supported by an overall ANOVA, using the $2 \times 2 \times 6$ factorial design of Figure 3. The effects of skewing interacted significantly with the number of stimuli, $F(1, 905) = 19.32, p < .001$; but the higher order interaction between skewing, stimuli, and categories was not statistically significant, $F(5, 905) = 1.29, p > .25$.

Figure 4 adds ratings for three pairs of 24-stimulus sets to those for five- and nine-stimulus sets (from Figure 3) that were rated using the same number of categories. Although there is a clear Category Effect with 24 stimuli, increasing the number of stimuli from 9 to 24 produced no additional Stimulus Effect. These conclusions are supported by a $3 \times 2 \times 3$ ANOVA following the design of Figure 4: The interaction between skewing and number of categories was again significant, $F(2, 566) = 23.46,$

$p < .001$, as was the interaction between number of stimuli and skewing (the Stimulus Effect), $F(2, 566) = 5.91, p < .01$. Planned comparisons showed that although the effects of skewing differed significantly for 5 versus either 9 or 24 stimuli ($p < .05$), the corresponding difference for 9 versus 24 stimuli was not significant ($p > .50$). Because subjects cannot reliably identify more than 9 stimuli that vary on a single dimension (Miller, 1956), corresponding 9- and 24-stimulus sets may not differ in the number and frequencies of their *effective* stimuli.

Fit of Simple Range-Frequency Model

The theoretical functions shown in Figures 5 and 6 represent the fit of the simple range-frequency model of Equation 1 to the data of Figures 2 and 3, with only the weighting parameter, w , left free to vary from panel to panel.

The frequency weighting for each panel, $1 - w$, was calculated by dividing the difference in ratings for positive and negative sets by the difference in frequency values for the two contexts. Algebraically, this method for calculating $1 - w$ is derived from Equation 1 as follows:

$$C_{i+} - C_{i-} = (wR_{i+} + (1 - w)F_{i+}) - (wR_{i-} + (1 - w)F_{i-}). \quad (3)$$

Assuming that $R_{i+} = R_{i-}$, range values drop out, so that

$$C_{i+} - C_{i-} = (1 - w)(F_{i+} - F_{i-}), \quad (4)$$

$$1 - w = (C_{i+} - C_{i-}) / (F_{i+} - F_{i-}), \quad (5)$$

where C_{i+} and C_{i-} are the mean category ratings of Squares 7 through 19 (using interpolated values when any of these stimuli were not in the set) for positively and negatively skewed distributions, and F_{i+} and F_{i-} are the frequency values of the same stimuli, calculated separately for positive and negative skewing. The estimate of $1 - w$ was based on just the middle stimuli (Squares 7 through 19) because with the two frequency values so close for end stimuli, very slight variations in ratings of the end stimuli would produce much greater variations in estimates of $1 - w$.

Calculation of frequency values used the counting algorithm employed in earlier applications of range-frequency theory (e.g., Parducci & Perrett, 1971). The total number of experimental presentations were first divided by the number of categories to obtain the number of presentations that would have been assigned to each category if subjects were simply following the frequency principle (while keeping a perfectly ordinal scale). For example, when a five-stimulus set of 50 presentations is rated using five categories, there would be 10 presentations in each category; thus for the positively skewed set, 10 of the 20 presentations of the smallest square would be rated *very small* and 10 would be rated *small* (for a mean rating of 1.5 on a 1-to-5 scale); 10 of the 14 presentations of the next largest size would be rated *medium* and 4 would be rated *large* (yielding a mean frequency value of 3.29); 6 of the 8 presentations of the middle size would also be rated *large* and the remaining 2, *very large* ($M = 4.25$); and the 4 presentations of each of the two largest squares would always be rated *very large* (5). For the negative set, the frequency values would be the mirror image of these five values (i.e., the smallest two sizes would have a fre-

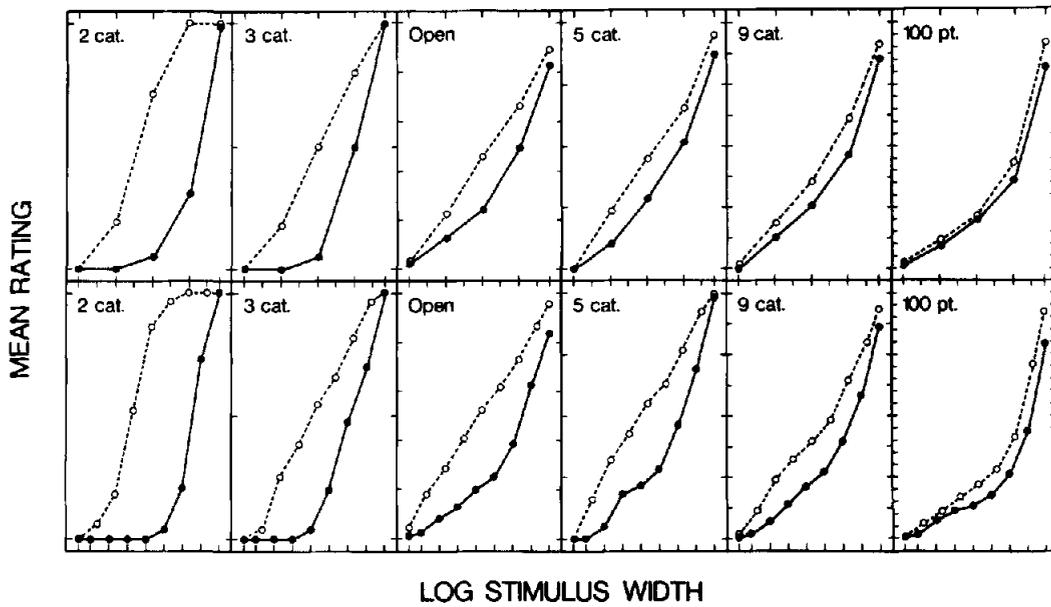


Figure 3. Stimulus Effect: effects of skewing greater for nine-stimulus sets (lower panels) than for five-stimulus sets (upper panels). (Category Effect similar for five and nine stimuli. Cat. = category; pt. = point.)

quency value of 1, the middle size would have a frequency value of 1.75, etc.). Separate sets of frequency values were calculated by this counting procedure for each panel, then substituted

along with the empirical category ratings into Equation 5, to obtain the empirically estimated values of $1 - w$ shown in the respective panels of Figures 5 and 6.

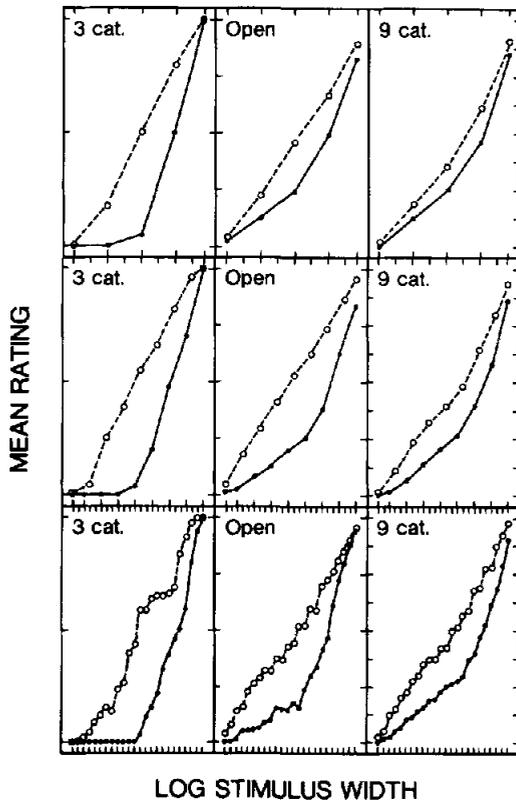


Figure 4. No additional Stimulus Effect for 24 stimuli (bottom panels). (Cat. = category.)

To arrive at the single range function that was used to fit all of the data (rightmost lower panel of Figure 6), range values for each rating function were first inferred by substituting the fitted value of $1 - w$ for that condition and the empirical ratings (linearly transformed to a 0-to-1 scale) into Equation 1 and solving for R_{ic} (separately for each stimulus). A single range function was computed by averaging the inferred range functions from nine-category, five- and nine-stimulus conditions. This function is close to the median of the range functions (after eliminating the highly irregular functions inferred for two categories) and is nearly linear to the widths of the squares—even though differences in Thurstonian discriminability are more logarithmic, as described in Parducci and Perrett (1971). This range function is similar in form to that typically found for squares (as compared with the log-linear relation typical of other psychophysical functions).

To predict the empirical ratings shown in any particular panel of Figure 5 or 6, range and frequency values for the stimuli were substituted into Equation 1 along with the $1 - w$ value fitted for that panel. Thus for all but the nine-category scales, the theoretical functions in each panel were fit by estimating only a single parameter, $1 - w$, from the data of that panel. The correlation between theoretical and data points was $r = .977$. The fit of the model is poorest for the two- and three-category scales. The two-category scales are really psychometric functions because mean ratings (on a scale from 0 to 1) are simply the proportion of *large* ratings; consequently they conform better to an earlier range-frequency model (described in Parducci, 1965) based on Thurstonian dispersions.

Although small but orderly skewing effects are shown by the 100-point scales, the extreme positive acceleration of these

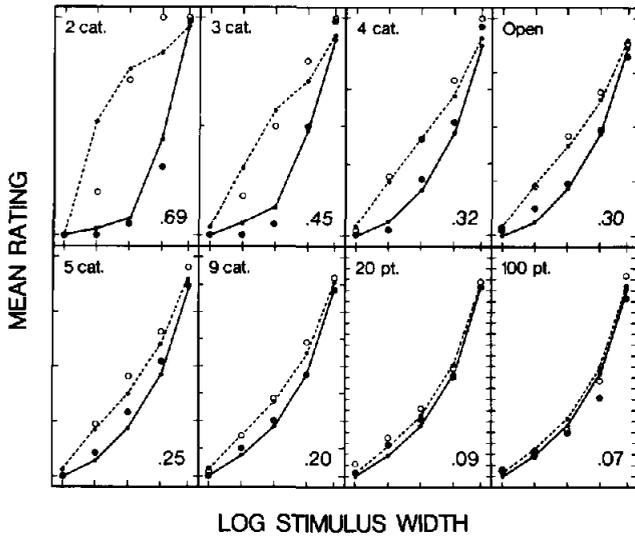


Figure 5. Theoretical fit (lines) for simple range-frequency model of Equation 1 to data from Figure 2. (Fitted value of single free parameter, $1 - w$, shown in each panel. Cat. = category; pt. = point.)

scales is typical of magnitude estimations when plotted against log stimulus width (cf. Parducci, 1963). Perhaps the 100-point scale may have more in common with magnitude estimations than with category ratings.

Attention to these anomalies should not detract from the major findings: (a) The effects of skewing decrease with increasing numbers of categories, the Category Effect; (b) the effects of skewing are greater for 9 than for 5 stimuli, the Stimulus Effect, but appear no greater for 24 than for 9 stimuli; and (c) both Category and Stimulus Effects can be described by changes in the single weighting parameter of range-frequency theory.

Fit of Retrieval-Consistency Model

The elaborated range-frequency model differs from the one just fitted in two essential ways: (a) The contextual frequencies used in calculating the frequency values for Equation 1 deviate from the actual frequencies of the presented stimuli, representing instead those frequencies retrieved from memory under the restrictions of the consistency rule; and (b) the value of w is independent of the number of categories or stimuli. Here it is assumed that range and frequency principles are weighted equally, that is, $w = .5$, a value close to those obtained from many of the previous fits of the simple range-frequency model.

The distribution of frequencies retrieved from memory is different for each condition. To determine the distribution of effective contextual frequencies for any particular condition, one must divide 12 by the number of categories (as explained in the introduction) to obtain the maximum number of presentations of each stimulus that enters into the calculation of frequency values. For example, the expected frequencies for a block of 12 presentations from the five-stimulus, positively skewed set would be close to 5-3-2-1-1. When rated using three categories, the depth of search is $12 \div 3 = 4$, so that the retrieved frequencies average close to 4-3-2-1-1. When rated using nine categories, the depth of search is $12 \div 9 = 1.33$, so that the

retrieved frequencies average close to 1.33-1.33-1.33-1-1. In the latter case the retrieved frequencies total to only six presentations so that the frequency principle would then assign an average of $6 \div 9 = .67$ of a presentation to each category.

Frequency values are calculated for each condition by applying the counting algorithm to these restricted sets, just as it was applied to the complete sets (the stimulus frequencies actually presented) in the fit of the simple range frequency model. With w set at .5, range values were inferred for each condition by substitution in Equation 1. The data were then predicted using a single range function, the median of the separately inferred range functions (after eliminating those for the two-category and 100-point scales, as in fitting the simple model). This overall range function is all that is estimated from the data, and it is assumed to be constant for all conditions. These common range values are then substituted back into Equation 1, along with the frequency values calculated for the "retrieved" distributions, to obtain the predicted category ratings shown in Figure 7.

Although the predictions are again terrible for two-category, positively skewed conditions (where a limen model would clearly be superior), they are almost as good as those fitted to the simple range-frequency model for the three-, open-, and five-category conditions. The predicted effects of skewing are too small for nine-category conditions, and this elaborated model (like the simple model of Equation 1) fails to capture the bowing of the 100-point scales—just as the overall range function precluded a good fit to 100-point data by the simpler model. What is encouraging is that with no empirical fitting of the skewing effects, these a priori estimates of what is retrieved from memory give a good account of both Category and Stimulus Effects. The retrieval-consistency model was not applied for 24-stimulus conditions because the additional parameters for determining how the identities of different stimuli might be confused would considerably complicate its application. The overall correlation between predicted and empirical data for Figure 7—as well as the five-stimulus, 4- and 20-point and nine-stimulus, seven-category scales (not shown)—is $r = .969$.

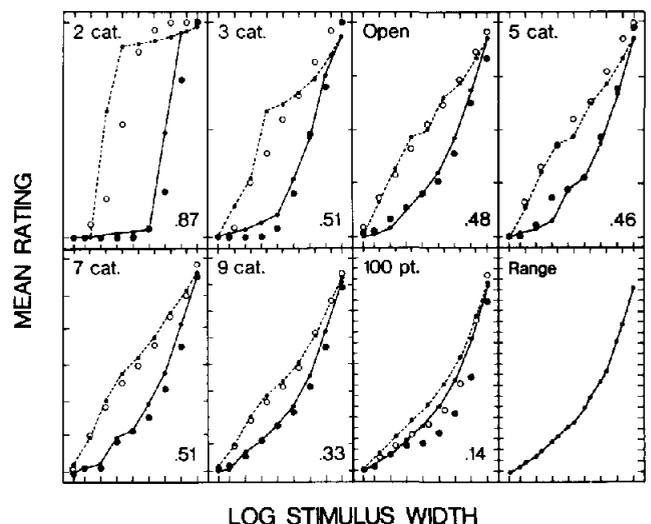


Figure 6. Range-frequency fit for nine-stimulus conditions. (Same range function [lower right panel] as for Figure 5. Cat. = category.)

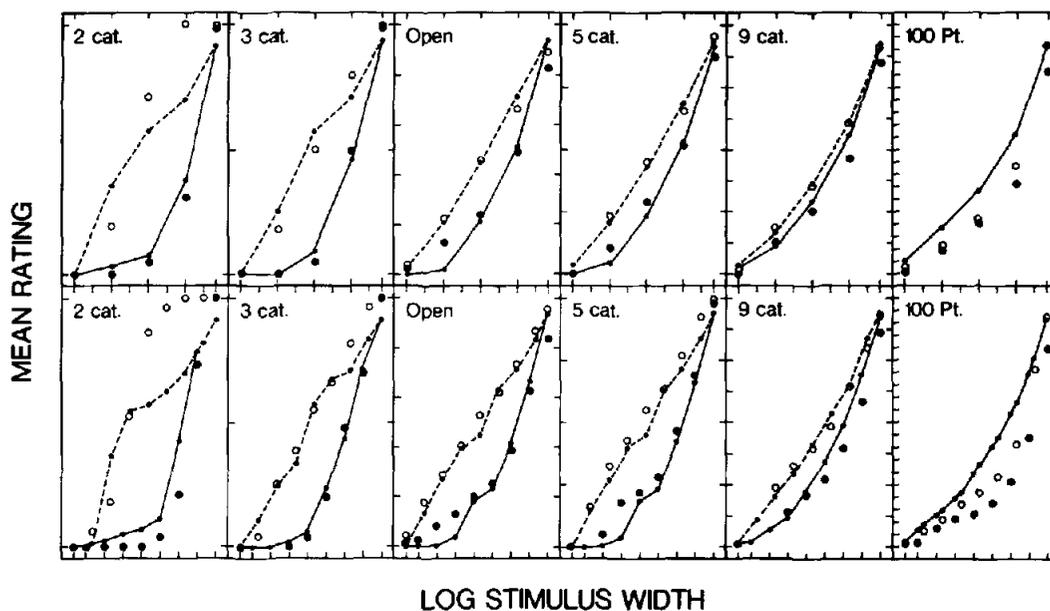


Figure 7. Fit of retrieval-consistency model to data of Figure 3, with effects of skewing for each panel predicted by consistency principle. (Cat. = category; pt. = point.)

Number and Maximum Frequency of Stimuli

The selection of stimulus frequencies for sets with different numbers of stimuli resulted in a natural confounding between the number of stimuli (5, 9, and 24) and how often the most frequent stimulus was presented per block of 50 (20, 12, and 6 times, respectively). Although these frequencies were selected to produce similar cumulative frequency functions, the confounding of number of stimuli with maximum stimulus frequency suggests a crucial test of the retrieval-consistency model. The derivation of the Stimulus Effect from the retrieval-consistency model depends on there being fewer repetitions of the more frequent stimuli when the total number of distinct stimuli is greater.

As a partial unconfounding, special nine-stimulus sets were constructed in which the frequency of the most frequent stimulus was doubled as shown in the bottom panels of Figure 8 (for comparison with the five- and seven-category scales elicited under standard instructions, top panels). Although the increase in maximum stimulus frequency actually increased the physical skewing (as measured by the area between the percentile rank functions), the differences in rating associated with skewing are smaller in the lower panels. This interaction between skewing and maximum stimulus frequency was statistically significant, $F(1, 253) = 11.81, p < .01$.

The elaborated range-frequency model predicts that the area separating the positive and negative functions would be reduced by 30%, close to the actual reduction of 36%. The reduction in the effects of skewing with greater stimulus frequencies adds considerable support for the retrieval-consistency elaboration of the simple range-frequency model.

Tests of Procedural Variables

In this section, we report the essentially null results of variations introduced to determine whether the Category Effect de-

pends upon restriction to prescribed categories, the particular labeling of the categories (either with different verbal labels or just with numerals), how the data are tabulated, or the time allotted for judgment.

Open scales. Two sets of open instructions provided either two example categories, Open(2), or six examples, Open(6). The Open(2) instructions stated: "If the presented square is Small in comparison with the other squares, record the letter S; if it seems Large, record an L. Feel free to make up whatever other categories seem appropriate, using as many or as few as you wish." For the Open(6) instructions, the following line was added after the *small* and *large* examples: "Use whatever categories seem appropriate: *Medium (M)*, *Slightly Smaller than Average (SSA)*, *Very Small (VS)*, *Very Very Large (VVL)*, etc. Use as few or as many categories as seems appropriate." After the presentation of experimental stimuli, subjects were asked to list the different categories they used in order from largest to smallest.

The procedure for tabulating responses for the open conditions assumes an underlying symmetry above and below the midpoint of each subject's scale: Corresponding to an overt *very very large*, there is assumed to be an unused *very very small*; and corresponding to an overt *slightly smaller than average*, there is assumed to be an unused *slightly larger than average*. For each subject, the number of categories inferred by assumption of underlying symmetry was used to determine the constants for the linear transformation to equate the endpoints of scales used by different subjects. For example, when *small*, *average*, *large*, and *very large* are the four categories actually used by a subject, the scale is assumed to include the missing *very small* so that *small* is tabulated as .25 on the 0-to-1 scale used for graphing. Although this procedure for normalizing the different open scales across subjects does rest upon arguable assumptions, it is easily and reliably applied.

The open scales shown in Figures 2 and 3 combine ratings

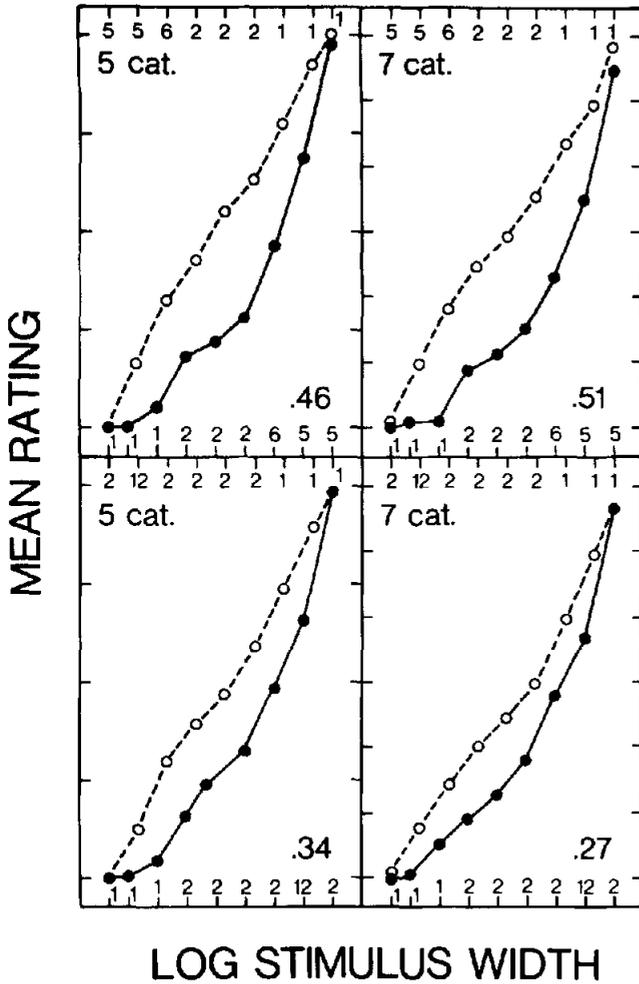


Figure 8. Effects of skewing reduced for modified nine-stimulus sets (lower panels); stimulus frequencies (for each block of 25) shown at top of each panel for positive sets, at bottom for negative sets. (Value of 1 = w shown in each panel. Cat. = category.)

from both sets of instructions. Figure 9 segregates these conditions by this instructional difference. The fit of the range-frequency model to these open scales is surprisingly accurate, supporting the method used for transforming the ratings.

Subjects in the Open(6) condition actually used about seven categories and those in the Open(2) averaged about five so that the greater effects of skewing for the latter is another example of the Category Effect. But it is of greater interest to determine whether the number of categories chosen spontaneously by the subject (i.e., apart from differences in instructions) produces a Category Effect. Figure 10 plots the transformed ratings for the five-stimulus, Open(2) conditions, segregating the 25 subjects who generated scales of five or more categories from the remaining 25. The empirical Category Effect is what would have been expected if the two subgroups had been instructed to use three and five categories, respectively (cf. the actual three- and five-category scales of Figure 2).

Verbal versus numerical recording. It does not appear to make any difference whether subjects record their verbal categories as letter abbreviations or as numerical ratings. Replica-

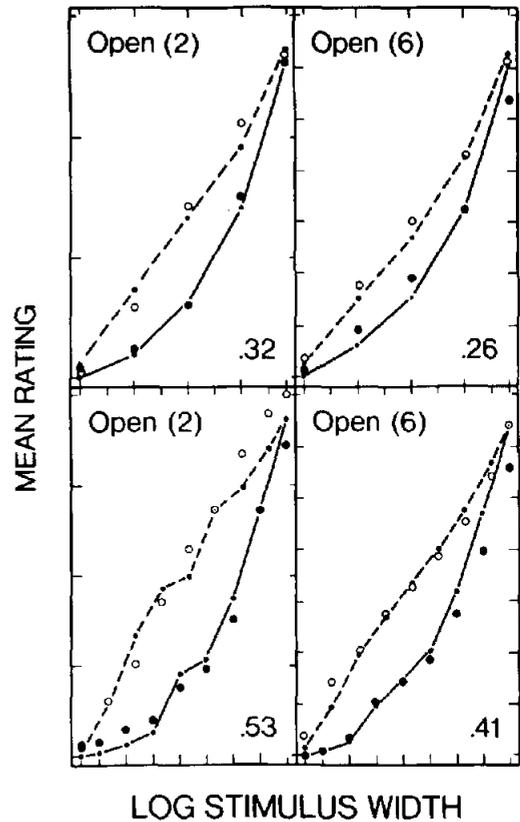


Figure 9. Category and Stimulus Effects for open scales: theoretical range-frequency fits for scales elicited by two examples (left panels) or six examples (right panels).

tions for three-, five-, and nine-category, five-stimulus conditions were run using numbers rather than the usual letter abbreviations. In both cases, response codes were listed with corresponding verbal categories at the top of the response sheets. The numerical recording produced scales that were virtually identical to those produced using the letter abbreviations

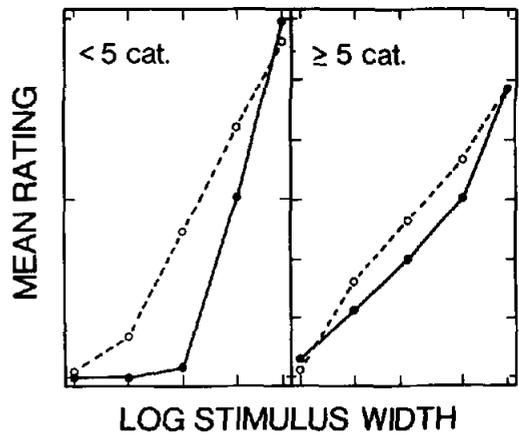


Figure 10. Category Effect when open (two-example) subjects use fewer than five categories (cat., left panel) versus five or more categories (right panel).

(consequently, the data were combined for Figure 2). The scales did change when subjects were instructed to *identify* each of the five stimuli with a number (1 through 5), and this manipulation reduced the effects of skewing ($1 - w$) by half. However, as in previous research (Tabachnick & Parducci, 1973), the effects of skewing were still significant.

Choice of verbal labels. Tabulating the category *large* on a three-category scale as equivalent to the category *very very large* on a nine-category scale may misrepresent underlying judgments that are possibly the same for both scales. To investigate the effect of verbal labels, two sets of instructions were used for seven-category, nine-stimulus conditions. In addition to the set described earlier (varying from *very very small* to *very very large*), a second set was constructed using the following category labels: *small*, *slightly larger than small*, *slightly smaller than average*, *average*, *slightly larger than average*, *slightly smaller than large*, and *large*. Categories and their letter abbreviations were printed at the top of response sheets, and subjects apparently had no problems using them. Although subjects were more likely to use the top category when it was *large* than when it was *very very large*, no other ratings differed by more than one standard error (.10 category-step). Because the effects of skewing were virtually the same for the two sets of categories, ratings from these two pairs of conditions were combined for Figure 6.

Method of tabulation. To evaluate the possibility that the method of tabulating categories numerically might account for the Category Effect, the nine-category, five-stimulus conditions were retabulated as though they represented only three categories. The three lowest categories (*very very small*, *very small*, and *small*) were each tabulated as 1 on a scale from 1 to 3; the next three categories as 2, and the three highest categories as 3. Although the retabulation did introduce some of the step-wise character of the regular three-category scales, the overall effects of skewing remained very small, with the value of $1 - w$ unchanged at .20 (as compared with the .45 obtained for the three-category scales).

As further evidence that Category and Stimulus Effects are not simply an artifact of how ratings are tabulated, Figure 11 illustrates these effects in terms of the stimulus rated *average*, that is, the adaptation level (after Helson, 1964). The advantage of using adaptation level as an *empirical* measure of the effects of skewing is that it requires no transformation of the data but simply a determination (by linear interpolation) of which stimulus value is rated *average*. To measure the effects of skewing, the adaptation level for the positively skewed set of stimuli was subtracted from the adaptation level for the negatively skewed set. This algebraic difference is plotted in the left panel of Figure 11. The difference in adaptation level decreases with more categories and fewer stimulus sizes. The value of the weighting parameter of range-frequency theory ($1 - w$) is presented in the right panel for comparison.

Development of the scale. The data from Experiment 1 have thus far been presented as means taken across all ratings in the regular series (viz., all 50 presentations). Figure 12 segregates the data for some of the five-stimulus conditions, presenting separate mean ratings for the first block of 10 trials (the preview), the second block of 10 trials, and the overall means across all 50 regular trials (taken from Figure 2). It is astonishing how quickly the scales develop: Within the 10-trial preview, the

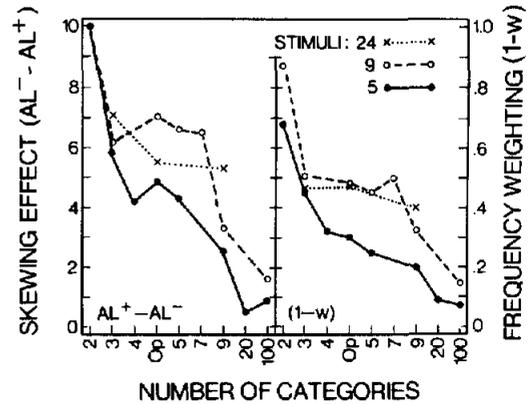


Figure 11. Effects of skewing on two measures: frequency weighting (right), differences in adaptation level (left). (Category Effect indicated by negative slopes, Stimulus Effect by higher values with more stimuli. $AL^- - AL^+$ represents difference in adaptation level between negatively and positively skewed sets.)

scales have already assumed the essential form that they will retain throughout the next 50 presentations. A very similar early development of the Category Effect is found for nine-stimulus conditions, with the Stimulus Effect also already present during the preview.

Whatever the process of apprehending the context for judgment, it does not take many trials. Although essentially complete information about the range and skewing of stimulus frequencies is available in the preview (which was constructed to be representative of the whole set), it is still surprising that so little repetition is required to establish the scale. It might have been more reasonable for subjects to assume that the skewing they were observing was a nonrepresentative sample from a uniform or perhaps a normal distribution of sizes. Instead, the data appear consistent with the "law of small numbers" (Tversky & Kahneman, 1974), in which people assume that a small random sample is more representative than the statistical laws of sampling would justify. That subjects are so quickly sensitive to contextual skewing is consistent with the assumption, basic to the retrieval-consistency elaboration, that the search set never contains more than 12 stimuli.

Time allocated for judgment. Is it possible that the Category Effect is simply the result of the limit on the time allocated for judgment by the experimenter? Bevan and Avant (1968) demonstrated that response latencies in a rating task increase with increases in the number of categories, which suggests that the subject is under greater time-pressure with more categories, resulting in reduced weighting of the frequency principle. To explore this possibility, additional subjects were run in the five-stimulus, nine-category conditions using presentation rates of either 2 s or 10 s. The resulting scales were virtually identical to those obtained using the regular 5-s interval. Thus, there is no evidence that the effects of skewing are influenced by the time allocated for judgment by the experimenter.

Summary Comments

The essentially null results of removing restrictions on which or how many categories are to be used, on whether numbers or

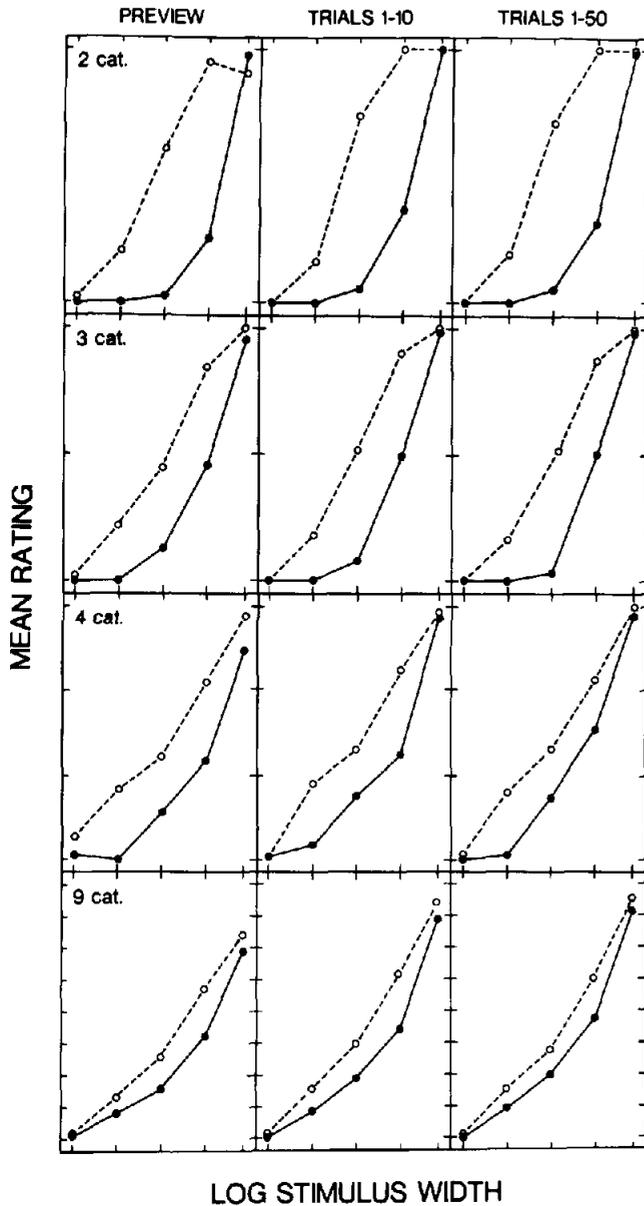


Figure 12. Early development of five-stimulus scales: little change between scales for preview, first 10 trials, and all 50 trials of regular series.

verbal labels are recorded, and on which particular verbal labels identify the categories all testify to the robustness of Category and Stimulus Effects. The insensitivity of the scales to large variations in the time allotted for judgment suggests that however elaborate the processes for determining range and frequency values are, they must occur very quickly, in less than a second. Although it is surprising that contextual effects are already apparent during the short preview, this seems consistent with the assumption that frequency values are based on not more than the preceding 12 presentations.

The success of the retrieval-consistency model in accounting for a large part of both Category and Stimulus Effects encourages further exploration of the conditions in which the hypothesized processes of judgment might be expected to occur. Experi-

ment 2 demonstrates that the Category Effect occurs even when the contextual stimuli are presented prior to instructions for judgment, that is, before the subject is told how many categories to use. Experiment 3 demonstrates the disappearance of Category and Stimulus Effects when stimulus frequencies are equalized, and hence conflict between frequency and consistency principles is removed. Experiment 4 shows that the findings of both Experiments 1 and 3 also hold for simultaneous presentation of stimuli, and the effects of its manipulation of stimulus arrangements lend additional support for the principle of consistency.

Experiment 2: Incidental Exposure to Contextual Stimuli

Although the success of the retrieval-consistency elaboration of range-frequency theory supports its interpretation of how the effective contextual skewing might be influenced by the number of categories, Experiment 1 forces little restriction on how this process might work. Is memory for past responses, in terms of specific categories, crucial to Category and Stimulus Effects? Haubensak (1982) reported that the frequency weighting for a three-category scale dropped to 0 when subjects were required to count backwards by threes during the 9-s interval between stimulus presentations. How is the retrieval of stimulus frequencies affected by conditions that might be expected to affect how stimuli are encoded for subsequent retrieval?

In Experiment 2 we attempt to answer these questions by independently manipulating instructions at two stages of the experimental session: (a) when the skewed sets are presented and (b) when ratings are later made of stimuli presented in a uniform distribution. Instructions for the first stage are assumed to affect how the skewed sets are encoded and stored for subsequent retrieval. Three tasks were used: In the incidental-exposure task (I), subjects were told only that they would later be tested on the stimuli presented (with no reference to judgment or to categories); in the successive comparison task (C1), each stimulus was to be compared with the immediately preceding stimulus (*viz.*, the presentation on Trial n with that on Trial $n - 1$); in the two-step comparison task (C2), each stimulus was to be compared with the second stimulus back (*viz.*, the presentation on Trial n with that on Trial $n - 2$). The three cover tasks were originally designed to test alternative theories of how encoding might affect memory from stimulus frequencies (see Wedell, 1984, for a discussion of these theories). After completing one of the three tasks, every subject in the second stage made conventional absolute ratings (as in Experiment 1) of a common set of test stimuli, using either three or nine categories.

If the Category Effect depends on stimuli being encoded and stored in terms of specified numbers of categories, no Category Effect should be found in the subsequent test: Subjects could not know when viewing the skewed sets how many categories they would later be instructed to use. On the other hand, if the number of categories does not affect how the stimuli are stored but only how they are retrieved for judgment, then the Category Effect should occur regardless of the cover task.

Method

Design

A four-way, Task \times Skewing \times Category \times Sequence, between-subjects factorial design was used. Task was manipulated at three levels: (I)

incidental, (C1) successive-comparison (compare one back), and (C2) two-step comparison (compare two back). There were two levels of each of the other three factors: skewing of task stimuli (positive and negative), category (three and nine), and sequence of the presentation and test stimuli (two block-randomized sequences each). An additional manipulation of size of the background field was subsequently introduced to throw light on anomalous data. Between 15 and 20 subjects participated in each condition, tested in groups of 8 to 12.

Stimuli

The stimulus sets were those used for the five-stimulus conditions of Experiment 1, as were the presentation sequences. Two test sequences were created by presenting each square twice, in randomized blocks of five (i.e., 10 test trials). Except for the added test sequences at the end, stimulus presentations were the same as in Experiment 1. In the reduced-background conditions, the lighted field was reduced from 80×120 cm to 45×60 cm by attaching a cardboard cutout to each slide so that the vertical height of the lighted field was only slightly larger than the largest square.

Procedure

The experiment consisted of two stages, task and testing. The 50 squares constituting the positive or negative set were presented in the task stage. All subjects were instructed to pay attention to each square and were warned that they would be tested afterwards (but they were not informed what the "test" would be). Subjects in the C1 and C2 conditions also received response sheets and additional instructions for one of the two tasks for comparative judgment. Their responses were restricted to just three categories: *larger*, *equal*, or *smaller*.

Immediately after completion of their respective tasks, all subjects received the same additional instructions and response sheets for the standard rating test. These instructions were similar to those of Experiment 1, but subjects were told to compare each test square with the whole set of squares that had been presented in the *first* part of the experiment. Ratings were to be made using either three or nine categories. The 10 test stimuli were then presented. Test Sequences 1 and 2 were paired, respectively, with Presentation Sequences 1 and 2 from Experiment 1.

Results and Discussion

Figure 13 shows systematically different results for the two sequences. For Sequence 1, the effects of skewing are consistently greater for three than for nine categories, and there is apparently no effect of task. The results from Sequence 2 are a mixed bag: Although there is a large category effect following the C2 task, it is virtually absent following either I or C1 tasks.

The four-way, between-subjects ANOVA again used the dependent variable of Experiment 1 (viz., the area below the rating function). Significant Skewing \times Sequence and Task \times Skewing \times Sequence interactions ($p < .05$) indicate that the effects of encoding are different for the two sequences. Consequently, separate three-way ANOVAs were conducted for each sequence. For Sequence 1, the Category \times Skewing interaction (the Category Effect) was significant, $F(1, 238) = 15.49, p < .001$, but the Comparison Task \times Skewing interaction was not, $p > .25$. The opposite was found for Sequence 2: The Comparison Task \times Skewing interaction was significant, $F(2, 205) = 8.09, p < .001$, but the Category \times Skewing interaction was not, $p > .25$.

The differences associated with sequence argue for opposed

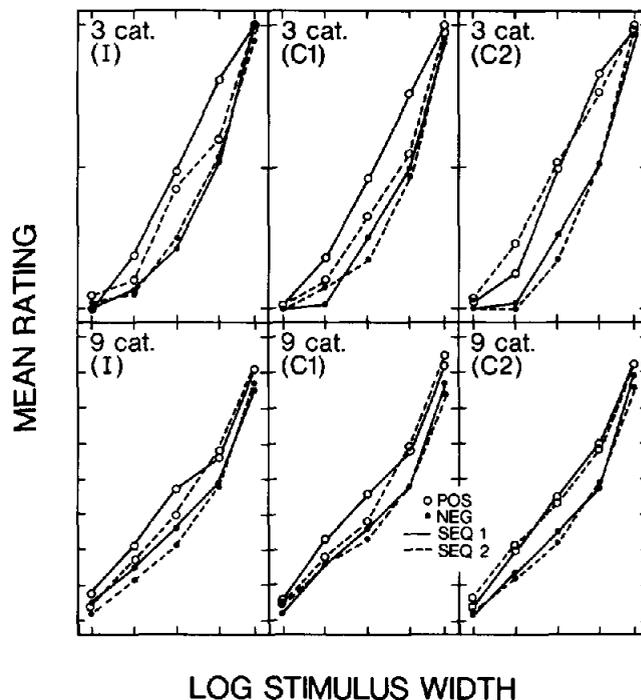


Figure 13. Category Effect for Sequence 1, three versus nine categories, but not for Sequence 2 with incidental (I) and compare-one-back (C1) tasks (Experiment 2).

conclusions about whether the category effect depends upon how stimuli are encoded. In an attempt to locate the critical differences between the two sequences, Newman-Keuls tests were used to analyze differences for each panel of Figure 13. The only significant interactions between skewing and sequence occur for the three-category, I and C1 conditions ($p < .05$). The effects of skewing with nine categories did not differ significantly for the two sequences. Thus, in order to interpret the results of Experiment 2, it is crucial to understand why the effects of skewing should depend on which random sequence was used in I and C1 conditions.

Although presentation and test sequences covaried, it would appear that it is the test sequence that is crucial, because no significant differences were found for the same presentation sequences in Experiment 1 or for other random presentation sequences in past research (e.g., Parducci & Perrett, 1971). For Sequence 1, the order of test presentations was 7, 19, 13, 1, 24, 1, 7, 24, 13, 19; for Sequence 2, it was 19, 13, 1, 7, 24, 7, 19, 24, 1, 13. The most important difference between the two may be in the initial presentation, Square 7 versus Square 19.

Figure 13 shows that Square 19 is rated much lower in Sequence 2 than in Sequence 1 for the three-category, positively skewed, I and C1 conditions (but not for the corresponding nine-category conditions). When Square 19 is presented first (in Sequence 2), it is consistently rated *average* on a three-category scale; however, when it is presented second (in Sequence 1), it is more often rated *large* than *average*. Because the simpler I and C1 tasks were designed to require less attention, subjects may have only partially reinstated the relevant stimulus content on the first trial of the test stage. Instead, they may have rated the first square in comparison with the greatest possible *range*

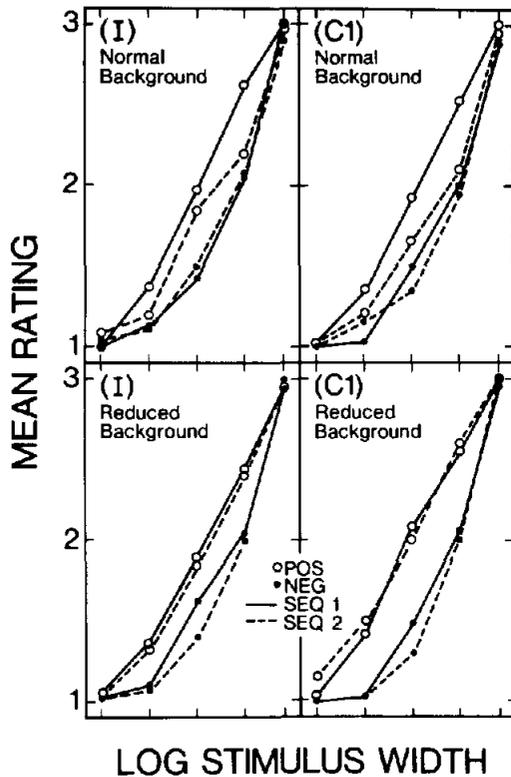


Figure 14. Elimination of sequence differences when height of background area is reduced to that of largest square (Experiment 2). (I = incidental task; C1 = compare-one-back task; pos = positive; neg = negative; seq = sequence.)

of sizes defined at its upper endpoint by the height of the lighted background. Subsequent ratings then tend to be consistent with the first rating. These considerations would not affect the already diminished skewing of the nine-category scales.

This account of the anomalous sequential effects entails that differences between the two sequences would be eliminated if the size of the background were reduced almost to the size of the largest square: When the upper endpoint of the range established by the background is the same as that of the stimulus set, it should make little difference which square is presented first. This was tested for the three-category, I and C1 conditions. Figure 14 compares results for normal- and reduced-backgrounds for these conditions. The large differences associated with the two sequences in the normal background largely disappear when the background is reduced. This reduction is demonstrated by a statistically significant three-way interaction between skewing, sequence, and background, $F(1, 298) = 7.93$, $p < .01$. Planned comparisons showed that although the skewing effect is significantly greater for Sequence 1 than 2 when the background is normal, there is no significant difference with the reduced-background condition for either I or C1 tasks. Thus, reducing the size of the lighted field largely eliminates the differences associated with sequence.

Because planned comparisons revealed a significant difference in skewing between normal- and reduced-background conditions only for Sequence 2, Sequence 1 may be considered to yield scales that are relatively independent of the back-

ground. The significant Category \times Skewing interaction for Sequence 1 encourages an interpretation of the Category Effect as independent of how stimuli are originally encoded for retrieval (e.g., retrieval of past responses). What establishes the Category Effect is the number of categories with which these stimuli are subsequently rated.

The absence of task effects for Sequence 1 (and for Sequence 2 with the small background) discourages speculation about how memory for stimulus frequencies might be encoded differentially for different numbers of categories. Indeed, there is substantial evidence suggesting that stimulus frequencies are encoded automatically, independent of encoding tasks (Hasher & Zacks, 1979). The absence of a task effect suggests that the search set includes a recent sample of presentations, made unrepresentative only by factors connected with the subsequent assignment of a specified set of categories for judgment.

It is a testimony to the generalizability of experimental research on category ratings that subjects can exhibit sensitivity to the skewing of sets exposed in comparison tasks that do not require category ratings or even, as in the incidental task, any particular type of evaluation at all. The results of these incidental exposure conditions also suggests that the Category Effect might be expected to occur more generally, for example, in survey research, where the researcher has no control over how stimulus events are initially encoded.

Experiment 3: Skewing the Spacing of Stimulus Values

According to the retrieval-consistency elaboration of the range-frequency model, frequency values become less sensitive to differences in stimulus frequencies as the number of categories increases. An important implication of this interpretation is that Category and Stimulus Effects should occur only when skewing is based upon differences in stimulus frequencies, not when skewing is based on differences in the spacing of stimulus values. When stimuli are presented with equal frequency, decreasing depth of search (i.e., the maximum number of repetitions counted for each stimulus) cannot make stimulus frequencies more equal; hence, frequency values will be independent of the number of categories or stimuli.

In earlier research (Parducci & Perrett, 1971), the simple range-frequency model provided impressive fits to a wide variety of stimulus spacings, including sets of squares bunched closely near the bottom of the range and others bunched closely near the top of the range; the effects of this kind of skewing did not depend upon the number of categories: The relationship between six- and nine-category scales was remarkably linear. However, that research was not designed to examine the Category Effect. The degree of skewing was smaller than for the sets of Experiment 1, and the difference in the effects of skewing for six versus nine categories would be expected to be small (cf. Figures 2 and 3).

In the present experiment, three- and nine-category and also 100-point scales are used. Although the degree of skewing corresponds to that used in Experiment 1, skewing is produced by varying the spacing rather than frequencies of the stimuli.

Method

Design and Subjects

A between-subjects, $3 \times 2 \times 2 \times 2$ factorial design was employed, with Category (3, 9 and 100), Skewing (positive and negative), Stimulus

(five and nine) and Sequence (both random) as independent factors. Approximately 10 undergraduates (sampled from the same source as Experiment 1) served in each condition, again in groups of 8 to 12 subjects.

Stimuli

The stimuli were drawn from the same pool of 24 squares used in Experiment 1. In the five-stimulus conditions, the positive set consisted of Squares 1, 3, 6, 11, and 24; the negative set consisted of Squares 1, 15, 20, 23, and 24 (these two sets were mirror images in terms of Thurstone discriminability scaling—cf. Parducci & Perrett, 1971). Stimuli were presented in one of two block-randomized sequences, with each stimulus occurring once in each block of five presentations for a total of 50 presentations. The large stimulus gaps for these sets (between Squares 11 and 24 for the positive set and between Squares 1 and 15 for the negative set) would require interpolation to infer the form of the rating functions over more than half of the stimulus range. To avoid this problem, two additional sizes were presented within the respective gaps, twice each, at the conclusion of the experimental session: these were Squares 15 and 19 for the positive set, Squares 7 and 11 for the negative set.

In the nine-stimulus conditions, the positive set consisted of Squares 1, 2, 3, 4, 6, 9, 12, 18, and 24; the negative set consisted of Squares 1, 8, 14, 17, 20, 21, 22, 23, and 24. This represented approximately the same degree of skewing (measured by the plots of mean percentile ranks against stimuli) as for the present five-stimulus sets (and the unequal-frequency sets of Experiment 1). Stimuli were presented in one of two, block-randomized sequences; each stimulus occurred once in each block of nine.

Procedure

The instructions and experimental procedure were virtually the same as for Experiment 1. The only differences were that for the present nine-stimulus conditions the preview consisted of 9 rather than 12 presentations, and the regular series of 54 rather than 50 presentations.

Results and Discussion

Figure 15 shows that the effects of skewing are nearly the same for three and nine categories, regardless of whether there are five or nine stimuli. This is in sharp contrast with Experiment 1 (frequency skewing, Figure 3) where a large Category Effect was obtained from three to nine categories. However, the value of $1 - w$ is markedly reduced for the 100-point scales for which the effects of skewing are actually larger for five than for nine stimuli. Although the overall Category \times Skewing interaction (Category Effect) is highly significant, $F(2, 222) = 9.66$, $p < .001$, the only significant ($p < .05$) differences are those separating the 100-point scales from either of the category scales for both five- and nine-stimulus conditions. The overall Stimulus Effect is not statistically significant, but planned comparisons reveal a significantly greater skewing effect with five stimuli for the 100-point scales (i.e., in the opposite direction from that in Experiment 1).

Predictions shown in Figure 15 were generated from the simple range-frequency model exactly as described in Experiment 1, using the same range function inferred for Experiment 1 and allowing only the frequency weighting to vary. Because each stimulus occurred equally often, the frequency values for simple and elaborated models are the same. However, if the elaborated model set w at .5, as in Experiment 1, it would overestimate the effects of skewing shown in Figure 15. In previous re-

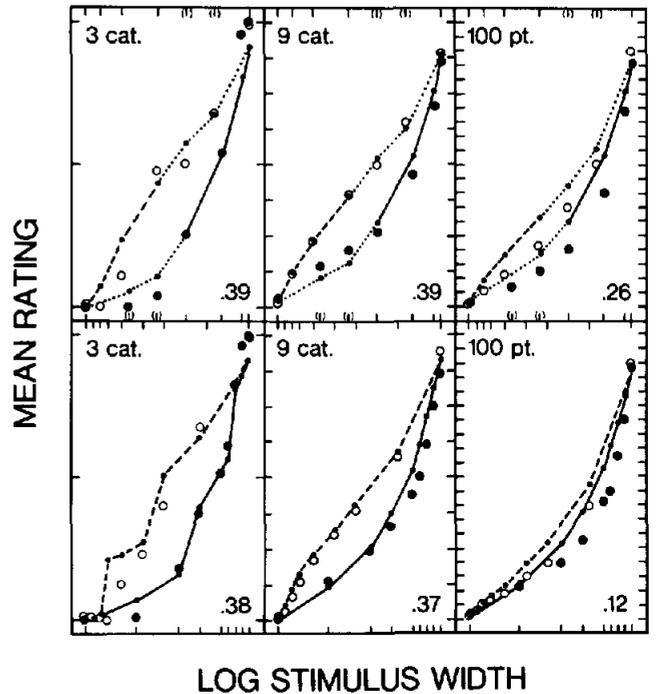


Figure 15. Successive presentations with skewed spacings (Experiment 3): no category effect for three versus nine categories (cat.); reversed stimulus effect for 100-point (pt.) scales. (Theoretical fits and values of $1 - w$ for simple range-frequency model of Equation 1.)

search (Parducci & Perrett, 1971), the best-fit weighting was close to .5 for less extreme manipulations of spacing. Perhaps the lower frequency weighting of the present experiment is due in part to a limit on the extent to which the scale of judgment may be shifted by extreme manipulations of stimulus spacing.

Although the nearly identical values of $1 - w$ for three- and nine-category, five- and nine-stimulus conditions provides strong support for the retrieval-consistency model, this simple picture is clouded by addition of the 100-point scales. The elaborated model did a poor job for the 100-point scales of Experiment 1 as well. Perhaps the 100-point scales are not just category scales with more categories.

Experiment 4A, 4B, and 4C employ the method of simultaneous presentation to answer questions about the generality of the findings with successive presentation and to further investigate the role of memory and of the consistency principle in judgment.

Experiment 4A: Simultaneous Presentation With Skewed Frequencies

Range-frequency theory was originally developed to account for category ratings of numerals presented simultaneously, that is, with the entire set printed on a single page. The goal was to specify the context for judgment as completely as possible so that the effects of different contexts could be more directly interpreted. Fits to data from experiments in which traditional psychophysical stimuli were presented successively encouraged application of the same range-frequency model. Although one might have expected that the representation of the context in

memory might deviate systematically from the context of stimuli actually presented, the Category and Stimulus Effects of Experiment 1 appear to be the first suggestion of systematic shifts in the effective skewing of contextual stimuli. Insofar as the reductions in the effects of skewing found in Experiment 1 reflect limitations on memory, one should not expect Category and Stimulus Effects to occur when the entire set of contextual stimuli can be apprehended at a glance.

In a recent study using *simultaneous* presentations (Mellers & Birnbaum, 1982), subjects rated dot patterns presented on a single page in either a positively or negatively skewed set. The effects of skewing did not differ much between 5- and 100-point scales; for both, the range-frequency weighting (w) was close to .5. Because Mellers and Birnbaum skewed stimulus *spacing* rather than frequencies, it is unclear whether the lack of a Category Effect is due to how the stimuli were presented or how they were skewed. Experiment 4A removes this source of confounding by presenting the skewed *frequencies* of Experiment 1 simultaneously on a single page..

Method

Design

Categories (3, 5, and 9), number of stimuli (5 and 9), and skewing (positive and negative) were the independent variables in a three-way, between-subjects factorial design. As in Experiment 1, skewing was produced by presenting the same small set of square sizes with varying frequencies.

Stimuli

The stimuli represented the five- and nine-stimulus sets of squares used in Experiment 1 (Table 2). There were 25 squares on a single, $8\frac{1}{2} \times 11$ -in. sheet of paper; widths of the squares were reduced to $\frac{1}{10}$ of their widths in Experiment 1. The page was divided into a 5×5 matrix of outlined rectangular blocks, with a square centered in each block. Squares were arranged randomly on the page in one of the random sequences from Experiment 1. A second, complementary arrangement was produced by simply inverting the page.

Instructions

The instructions were also very similar to those of Experiment 1, except that each verbal category was also identified by its rank in the set of categories (e.g., $1 = \text{small}$). Subjects were instructed to judge how large or small each square appeared relative to all the other squares and to record the appropriate numerical label within the rectangle surrounding the square being rated.

Subjects

Subjects were 240 undergraduates, tested either in one group of 54 as part of a regular class lecture or in groups of 8–12 in a separate laboratory. Data from 2 subjects who neglected to rate all of the stimuli were not used.

Results and Discussion

Figure 16 presents clear evidence for a Category Effect with simultaneous presentations. As in Experiment 1, there is also a Stimulus Effect. The procedure for fitting the simple range-frequency model was the same as described in Experiment 1,

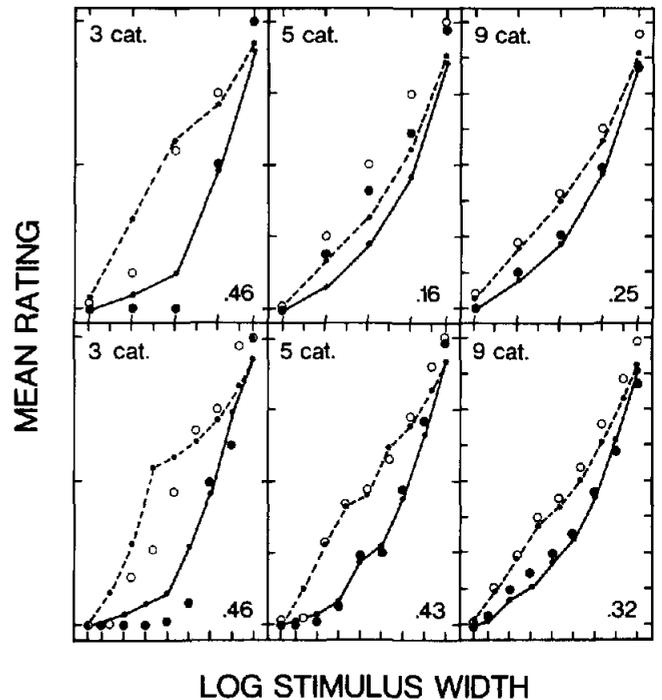


Figure 16. Simultaneous presentation: Category (cat.) and Stimulus Effects with frequency skewing of squares (Experiment 4A). (Theoretical fit and values of $1 - w$ for range-frequency model of Equation 1.)

and the same range function was used. The theoretical fits and corresponding values of $1 - w$ are similar to those obtained with successive presentation (cf. Figures 5 and 6).

A three-way ANOVA (Category \times Stimulus \times Skewing) reveals a statistically significant category effect (viz., the Category \times Skewing interaction), $F(2, 226) = 9.83, p < .001$. The Stimulus Effect is also significant, $F(1, 226) = 4.09, p < .05$, but the three-way interaction is not, $.10 < p > .05$.

These results are very similar to those for successive presentation with unequal frequencies (Experiment 1) and thus suggest that the lack of a Category Effect in the Mellers and Birnbaum (1982) study was due to the use of skewed spacing rather than to the stimuli being presented simultaneously. The presence of Category and Stimulus Effects even when the contextual skewing can be apprehended at a glance casts doubt upon any interpretation of these effects based strictly on differences in memory.

On the other hand, the retrieval-consistency model would predict Category and Stimulus Effects for simultaneous presentation, which could hardly be expected to reduce the importance of consistency. The five-category scales for the present five-stimulus conditions add special support to the notion that subjects follow the principle of consistency. These two scales are much more linear than the corresponding scales of Experiment 1 (or than the range-frequency fits). In this special case, it must have been obvious to most subjects that the number of categories corresponds exactly to the number of different physical sizes, and it was relatively easy to maintain consistency by using the five categories simply to rank the five sizes.

Experiment 4B: Ordered Arrangements

In earlier work with simultaneous presentations (Parducci & Marshall, 1961a), the relative weighting of range and frequency principles was dramatically affected by how the stimuli were arranged: The effects of skewing were greater when the experimental arrangement made it easier to apprehend the relative frequencies of stimuli in different parts of the range. Similarly, Parducci & Marshall (1961b) found greater effects of skewing for numerals when they were arranged ordinally on a page than when they were presented successively in random order. These findings suggested that the range-frequency weighting (w) may vary with the difficulty of applying the frequency principle. Although the number of categories was not varied in the earlier research, it is possible that category and stimulus effects might result from increased difficulty in applying the frequency principle with more categories or fewer stimuli.

Experiment 4B tests this possibility by arranging the stimuli on the page so as to make the skewing of frequencies more immediately apparent, and hence application of the frequency principle easier. Interpretations of Category and Stimulus Effects based upon difficulty of retrieval or integration of frequency values imply that such arrangements should reduce Category and Stimulus Effects. On the other hand, if category and stimulus effects result from a conflict between frequency and consistency principles, ordering the stimuli should not reduce these effects but may even enhance them insofar as the ordering also draws attention to the number of different stimulus values.

Method

Design

In addition to the squares employed in Experiment 4A, numbers and lines were also used. Numerals facilitate generalization beyond psychophysics to an abstract domain that has been studied intensively but without previous manipulation of the number of categories (e.g., Parducci et al., 1960). The lines were selected because they provide a convenient graphic representation of the skewing of lengths. The experiment consisted of three separate factorial designs, one for each type of stimulus. There were three between-subjects factors: Number of Categories (three and nine), Skewing (positive and negative), and Number of Stimulus Groups (five and nine, except for lines which consisted of either five different line lengths or five different groups of lines with very small variations in the lengths within each group). The positively and negatively skewed sets for each type of stimulus were created by varying the relative frequencies of a small set of stimuli or stimulus groups. To prevent confusion between instructions and also to minimize interference from previous sets, each subject was exposed to the same direction of skewing and number of stimuli for all three types of stimuli and was also instructed to use the same number of categories throughout. Order of presentation of the three types of stimuli was counterbalanced so that each type was rated equally often in first, second, and third positions. The 25 stimuli of each condition were arranged on the page in order of magnitude.

Stimuli

Squares. These were the same as used in Experiment 4A (five- and nine-stimulus, positively and negatively skewed, unequal-frequency sets); the only difference was that the squares were arranged on the page

in ascending order (left to right, top to bottom) rather than random order.

Numerals. These varied from 108 to 992. Skewing was achieved by varying the relative frequencies of stimulus groups. These frequencies were the same as those used for the squares in Experiments 1 and 4A. To avoid repeating the very same numeral, repetitions of a group consisted of numerals of similar value: Every numeral in a group was within 1 unit of another numeral in the same group; for example, the positively skewed, five-group set consisted of the following numerals: 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 326, 327, 328, 329, 330, 331, 332, 548, 549, 550, 551, 770, 771, 991, and 992. The 25 numerals in each set were arranged in a single column on the page, from smallest to largest. For conditions with five groups of stimuli, numerals 108, 329, 550, 771, 881, and 992 were common to positive and negative sets; for those with nine groups of stimuli, numerals 108, 219, 329, 439, 550, 661, 771, and 992 were common to both sets. The ratings of these common stimuli were used to assess the effects of skewing.

Lines. These varied from 9 to 161 mm, with a width of approximately 1 mm. In the equal-length conditions, skewing was achieved by varying the frequencies of the five lines, 9, 47, 85, 123, and 161 mm (which were equally spaced in terms of physical length). In the unequal-length conditions, no line was repeated exactly; instead, repetitions consisted of adding lines of very similar value (every line in a group was within 1 mm of another line in the same group). The relative frequencies of these groups were the same as for the five-stimulus squares and numerals. The 25 lines in each set were arranged in descending order on the page, with the right margin (35 mm) the same for each line. Ordered arrangements of lines made the skewing of frequencies particularly apparent.

Instructions

General instructions were printed on the first page of the experimental booklet. These emphasized that subjects were to work through the booklet, from front to back, completing each page before going on to the next. Specific instructions for each type of stimulus were printed on the page directly preceding the set to be judged. These instructed subjects to rate either squares, lines, or numerals, using either a three- or nine-category scale. Each category had the verbal label used in Experiment 1, for example, *small*, *average*, *large*, for the three-category scales, except that *short* was substituted for *small*, and *long* submitted for *large* for ratings of lengths. Again, each verbal label was identified by a letter abbreviation, for example, *VVS* for *very, very small*. Subjects were to record the abbreviation corresponding to their judgment (either within the rectangle surrounding each square, to the right of each line, or to the left of each numeral).

Subjects

Subjects were tested in a group of 226 during their regular class lecture with one of the authors. Ratings of squares from 2 subjects and of numerals from 10 subjects were eliminated either because of a failure to maintain an ordinal scale or a failure to use at least half the range of available categories.

Results and Discussion

Figure 17 presents mean ratings for squares, numerals, and lines, respectively. In each case, there is clear evidence for both Category and Stimulus Effects. The ratings of the squares provide the most direct comparison with previous results. The pattern of results is very close to that obtained using either successive presentation (Experiment 1, Figures 2 and 3) or simultaneous presentation (Experiment 4A, Figure 16) of these same sets of squares but with random sequences or arrangements.

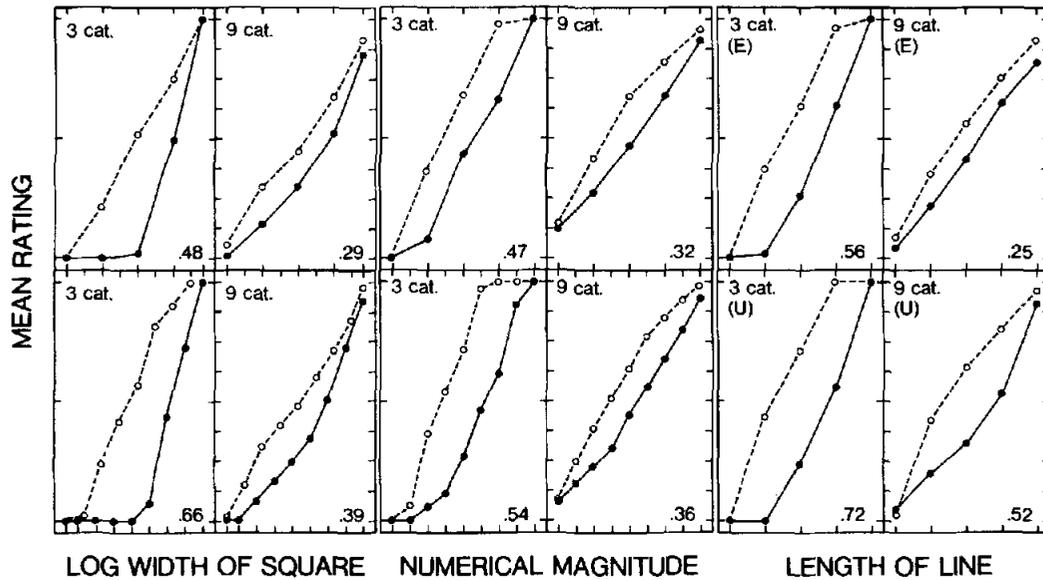


Figure 17. Simultaneous presentation of ordered squares, numerals, and lines (Experiment 4B): Category (cat.) and Stimulus Effects with frequency skewing. (Stimulus Effect for lines reflects homogeneity of lengths within each group—either equal [E] or unequal [U].)

Parallel results were obtained for ratings of ordered numerals and lines when lengths within each group are equal. Allowing the line lengths within each group to vary slightly produces a clear Stimulus Effect.

A separate three-way ANOVA was conducted for each type of stimulus, with the area beneath the rating function again used as the dependent variable. The Category Effect was statistically significant for all three types of stimuli, $p < .001$. The Stimulus Effect was significant for numerals, $F(1, 208) = 6.36, p < .05$, and for lines, $F(1, 218) = 9.85, p < .01$, but not for squares, $F(1, 216) = 1.71, p > .10$.

The demonstrations that large Category and Stimulus Effects are obtained even when simultaneously presented stimuli are arranged so that their respective frequencies are made more apparent further reduces the credibility of any interpretation of these effects predicated upon difficulty of taking stimulus frequencies into account. Instead, these data lend further support to the principle of consistency. This principle conflicts with the frequency principle when nine categories are used to judge just five different groups of stimuli occurring with unequal frequencies. Subjects resolve the conflict by favoring consistency, rarely assigning the same stimulus to more than one category.

Increasing the number of stimulus groups enables the subject to identify them more precisely (i.e., transmit more information) by using more categories. This is clearest for lines rated using nine categories. When all lengths are equal within each group, all receive the same rating, and the effects of skewing are small. But when lengths vary slightly within each group, most subjects use more than one category to represent the different lines of the most frequent group (with dramatic increase in $1 - w$).

Experiment 4C: Simultaneous Presentation With Skewed Spacing

Experiments 1 and 3 support the empirical generalization that Category and Stimulus Effects occur with successive pre-

sentations only when stimulus frequencies are skewed, not when stimulus spacing is skewed (at least not for three vs. nine categories). This generalization also appears to hold for simultaneous presentations: Experiments 4A and 4B demonstrate Category and Stimulus Effects by using unequal frequencies, but Mellers and Birnbaum (1982) did not find a Category Effect by using unequal spacing. Experiment 4C attempts to replicate Mellers and Birnbaum (1982), also using simultaneous presentation of other types of stimuli to test for the Category Effect with skewed spacing.

Method

Design

In addition to the dot patterns employed by Mellers and Birnbaum (1982), squares and numbers were also used. Squares facilitate comparison with Experiments 1, 3, 4A, and 4B, numerals with Experiment 4B.

Experiment 4C employed three separate factorial designs, one for each type of stimulus. There were three between-subject factors: Categories (3, 5, 9, and 100), Skewing (positive and negative), and Arrangement (two stimulus arrangements, both random—except for the number-judging task in which one arrangement was in order of increasing magnitude). As in Experiment 4B, each subject was given the same number of categories and the same direction of skewing for all three types of stimuli, and order of presentation of the three stimulus sets was again counterbalanced.

Stimuli

Dot patterns. The sets of dot patterns were those used by Mellers and Birnbaum (1982). Each pattern consisted of solid black dots, 1 mm in diameter, scattered irregularly within a square, 25 mm on a side. Eleven of these patterns were printed together on a single, $8\frac{1}{2} \times 11$ -in page. Six dot patterns (with 12, 18, 27, 40, 60, and 90 dots, respectively) were common to both sets; their ratings were used to assess the effects of skewing. For the positive set, five low-density patterns were added (with 14, 15, 16, 21, and 23 dots, respectively). For the negative set, five high-

density patterns were added (with 47, 51, 70, 74, and 77 dots, respectively). Because each pattern occurred only once, skewing was manipulated solely by variation in the spacing of stimulus values. One of the two random arrangements was identical to that used by Mellers and Birnbaum. Subjects recorded each rating above the pattern being judged.

Squares. The squares were proportional in size to Squares 1 through 11 from Experiment 1. They were outlined rather than solid, with the largest being 5.7 cm in width. Squares 1, 3, 5, 7, 9 and 11 were presented once each in both sets, and their ratings were used to assess the effects of skewing. To make the positive set, four repetitions of Square 2 and two of Square 4 were added. For the negative set, four repetitions of Square 10 and two of Square 8 were added. Thus for the squares, skewing was produced by manipulating both frequency and spacing (each contributing approximately equally to the skewing). Subjects recorded their ratings inside the squares.

Numerals. These were similar to those used earlier by Parducci et al. (1960). Forty-four numerals, varying in magnitude from 108 to 902, were typed in four columns of 11 each. Ratings of the 14 numerals common to both positive and negative sets, namely 108, 169, 230, 291, 352, 413, 474, 535, 596, 657, 718, 779, 840, and 902, were used to assess the effects of skewing. For the positive set, skewing was achieved by adding 26 numerals below 505 (the midpoint) and four above it. For the negative set, 26 numerals were added above the midpoint and four below it. No numeral appeared more than once on the page. One arrangement was in order from smallest to largest, starting at the top of the leftmost column and ending at the bottom of the rightmost column. The order of the other arrangement was random. Subjects recorded their ratings to the right of each numeral.

Instructions

General instructions, printed on the first page of the experimental booklet were the same as those for Experiment 4B. Specific instructions for each type of stimulus were printed on the page directly preceding the set to be judged. These instructed subjects to rate either the squares, dot patterns, or numerals, using either a 3-, 5-, 9-, or 100-point scale. Only the two ends of the rating scales had verbal labels: The lower end was labeled 1 and either *very small* (for squares and numerals) or *very light* (for dot patterns) and the upper end was labeled either 3, 5, 9, or 100 and either *very large* or *very dark*. Subjects were also instructed to anchor their scales by letting the lowest rating correspond to either the smallest square, lightest dot pattern, or smallest numeral on the page and the highest rating (3, 5, 9, or 100) to either the largest square, darkest dot pattern, or largest numeral on the page. This end-anchoring followed the procedure employed by Mellers and Birnbaum (1982); the results of pilot research conducted to account for their not getting the Category Effect suggested that neither the introduction of end-anchoring nor recording by numerals was crucial.

Subjects

Subjects were tested either in groups of 35, 85, or 150 during their regular class lecture or in groups of 8 to 12 in a separate laboratory. Data from almost one quarter of the 317 subjects were not used either for failing to complete all the judgments (time pressure allowed only 15 min) or for not following instructions. Nearly all of the eliminated subjects had been tested in their classroom lecture hall.

Results and Discussion

Figure 18 presents the data of Experiment 4C. The category effect is virtually absent for each of the three stimulus dimensions. In spite of some variability in the effects of skewing, nowhere do we find the systematic decline with increased numbers

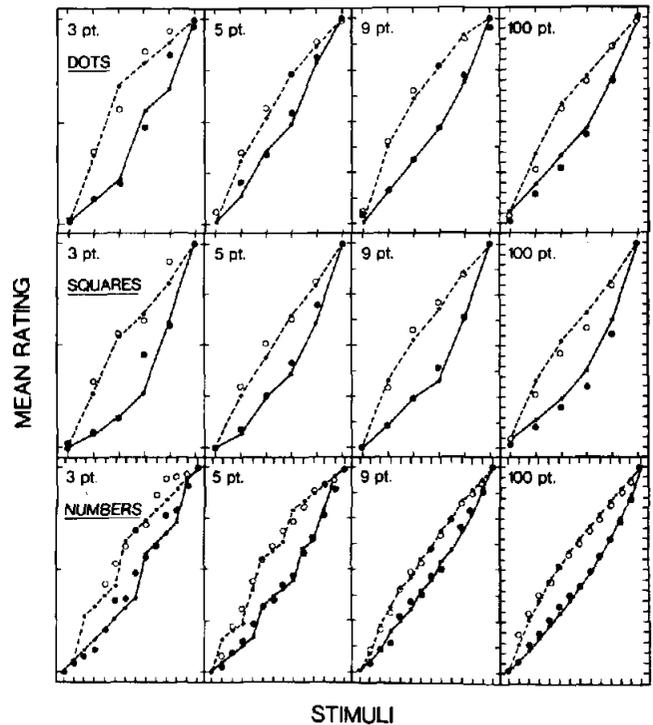


Figure 18. Simultaneous presentation: no Category Effect for any of three types of stimulus (Experiment 4C). (Theoretical fits for simple range-frequency model of Equation 1 [using weightings from Table 3]. Pt. = point.)

of categories. Even the 100-point scales show large systematic effects of skewing, and unlike Experiments 1 and 3, they are fit well by the simple range-frequency model (thus appearing closer to category ratings than to magnitude estimations). Although it seems inappropriate to apply the retrieval-consistency model when so few stimuli (12) are presented all together, one might have expected perhaps a small Category Effect for squares because skewing was determined equally by differences in spacing and frequencies.

Three-way (Category \times Skewing \times Arrangement) ANOVAs were run separately for each type of stimulus because the differences between their frequency distributions of stimuli confound direct comparison. There was a significant interaction between category and skewing for dot patterns, $F(3, 224) = 3.02, p < .05$ but not for squares or numerals, $p > .10$. As shown in Figure 18, the effects of skewing were especially small for the 5-point scale for dots: A Newman Keuls post hoc test (at the $p < .05$ level) revealed that they were significantly smaller than those for three- and nine-point dot conditions. But this irregular interaction is *not* the systematic decline (Category Effect) found when stimulus frequencies are skewed.

The values of $1 - w$ presented in Table 3 and the fits of the simple range-frequency model were determined by using Equation 5 over a broad subrange for each stimulus dimension (Squares 3-9, Dots 18-60, or Numbers 169-840). A separate range function was determined for each of the three types of stimuli; because inferred range values were fairly linear to numerical magnitude, square width, and log number of dots, lin-

Table 3
Frequency Weightings (1 - w) for Different Scales and Arrangements of Experiment 4C

Scales (points)	Dot patterns			Squares			Numbers		
	R-1 ^a	R-2	M	R-3	R-4	M	R-5	Ord. ^b	M
3	.51	.39	.45	.32	.45	.39	.13	.53	.33
5	.25	.39	.32	.44	.32	.38	.29	.63	.46
9	.73	.47	.60	.61	.51	.56	.09	.51	.30
100	.67	.45	.56	.55	.57	.56	.24	.72	.48
M	.54	.43	.48	.48	.46	.47	.19	.60	.39

^a R-1 through R-5, five random arrangements. ^b Arranged in order of magnitude. Ord. = ordered.

ear regression was used to derive the best-fitting range function for each dimension.

One anomaly of the data is the extremely low values of $1 - w$ for the randomly ordered arrangement of numerals: The overall mean frequency weighting is only .19 (as compared with .60 for the ordinal arrangement); this interaction between skewing and arrangement of numerals was statistically significant, $F(1, 224) = 49.97, p < .001$. The reduced effects of skewing for random versus ordered arrangements of numerals is consistent with previous research (Parducci & Marshall, 1961a, 1961b), suggesting that w may vary with the difficulty of apprehending frequency values, although such difficulty is hard to quantify.

The data for Experiment 4C are generally supportive of the retrieval-consistency elaboration of the range-frequency model, which entails the elimination of the Category Effect when stimulus spacing, rather than frequencies, is skewed. The disappearance of the Category Effect should also put to rest any lingering concern that the Category Effect could be some artifact of how the ratings are normalized by transformation. The same tabulation procedures and linear transformations (Equation 2) used for Experiment 1 were also used here and for Experiment 3, where spacing does not produce the category effect found for frequency skewing.

General Discussion

Experiment 1 demonstrated that the effects of skewing the stimulus frequencies vary inversely with the number of categories, directly with the number of stimuli. Both Category and Stimulus Effects were also found when subjects were encouraged to devise their own categories of judgment (i.e., with open scales). The essentially null results of various procedural manipulations attest to the robustness of these effects: The obtained scales were independent of the verbal labels defining the different categories, the method of recording ratings (numerals or letter abbreviations), the different methods of tabulating the ratings (individually or grouped by threes), and the time allotted for judgment (2, 5, or 10 s).

In Experiment 2, the Category Effect was also found following incidental exposure to skewed distributions of stimuli, exposure in which subjects could not have known how many categories would later become available for judgment. These results suggest that the Category Effect could also occur in survey research where the researcher has no control over how contextual events are initially encoded. The robustness of Category and

Stimulus Effects was demonstrated further under the conditions of simultaneous presentation employed in Experiments 4A and 4B.

However, neither Category nor Stimulus Effects were found when skewing of the stimulus sets was achieved by varying the spacing, rather than the frequencies, of stimulus values (Experiments 3 and 4C). The combined results of all our research support the empirical generalization that Category and Stimulus Effects occur only when stimuli are presented with unequal frequencies.

The obtained shifts in the effects of skewing can be characterized with precision by changes in a single empirically determined value, w , the weighting parameter of the simple range-frequency model (Equation 1). Formally, w represents the relative weight subjects assign to range versus frequency principles: Subjects give less weight to equalizing the division of stimulus frequencies among categories when there are either more categories or fewer stimuli. In past research (e.g., Parducci & Marshall, 1961a, 1961b), w was hypothesized to reflect the relative difficulty of applying the frequency principle. However, in the absence of an independent measure of difficulty, this hypothesis is not a very useful predictor of the effects of different experimental conditions (e.g., unequal frequencies vs. unequal spacing). A more promising interpretation of changes in w incorporates considerations of consistency, the principle that only one category is assigned to each stimulus.

The retrieval-consistency elaboration of range-frequency theory attributes Category and Stimulus Effects to systematic changes in the effective frequency values. Whereas the fit of the simple range-frequency model of Equation 1 requires separate estimates of w for each change in the number of categories or stimuli, the elaborated model *entails* the effects of these changes and thus increases the predictive power of range-frequency theory. Because the model attributes the Stimulus Effect to the *confounding* of changes in stimulus frequencies rather than to changes in the number of different stimuli, it correctly predicts the reduction of skewing effect when the same physical skewing is achieved by an increase in the frequency of presentation of the most frequent stimulus (Experiment 1). This same model also implies the disappearance of Category and Stimulus Effects when skewing depends upon the spacing of stimulus values presented with equal frequency. Although the implications of the elaborated model are less clear when all contextual stimuli are presented simultaneously, the effects of how these stimuli are

arranged were, in general, supportive of the consistency principle, which is the basis for the retrieval-consistency elaboration. Perhaps the most striking example of subjects following this principle was the almost complete consistency with which squares of the same size were rated when five sizes were presented simultaneously and rated with five categories.

An Information Processing Interpretation

Earlier applications of range-frequency theory have assumed that subjects base their judgments on the entire set of stimulus presentations. This assumption was supported by the encouraging fits of the simple model to data from experiments in which psychophysical stimuli were presented successively, in long randomized series (e.g., Parducci, 1963, 1965; Parducci & Perrett, 1971). The discovery of the Category Effect stimulated speculation about how using more categories could reduce the effective skewing. The results of the present experiments suggest that subjects systematically discount differences in stimulus frequencies when more categories are available, as though the depth of search, or maximum count of past frequencies, varied inversely with the number of categories. The principle of consistency provides a basis for inferring the depth of search directly from the number of categories.

The principle of consistency seems intuitively appealing. As a subject in such an experiment, one resists applying more than a single category to the same stimulus. When ordered simultaneous presentations make any such switching of categories obvious (Experiment 4B), all repetitions of the same stimulus are placed in the same category. Inconsistent assignment of stimuli is less obvious to the subject with successive presentations, and any role of the consistency principle in the calculation of frequency values would have to be an unconscious part of the judgment process.

The retrieval-consistency model assumes that the principle of consistency enters the judgment process at the stage where past presentations are retrieved from the search set and transferred to working memory, which is conceived as the effective context for judgment. The assumption that the search set is limited to the last 12 stimulus representations is supported by transfer experiments (Wedell, 1984) in which presenting 20 or more rather than 10 presentations of squares from an initial skewed set had little additional effect on the rate of adaptation to a second set skewed in the opposite direction. Although earlier presentations do not get transferred to working memory, the two end-stimuli apparently remain. If these ever dropped out, range values would shift immediately, resulting in a sudden shift in the scale. The retention of the two end-stimuli in the effective context is supported by earlier research on extension versus restriction of the stimulus range (Parducci, 1956): When the range was extended, the scale shifted quickly and dramatically, but when the range was restricted, the scale shifted very little—as if subjects retained the end-stimuli of the initial set in memory.

In the transfer of stimuli to working memory, the principle of consistency limits the number of stimulus repetitions to the frequency of use of each category as dictated by the frequency principle (i.e., number of stimuli in the search set divided by number of categories). This limit on the count of stimulus repetitions eliminates conflict between frequency and consistency

principles. Thus, the stimuli in working memory constitute the effective context for judgment, that is, the context that is used to calculate range and frequency values. The judgment of the presented stimulus is expressed as the weighted average of range and frequency values for this effective context, in accordance with Equation 1. The value of w was fixed at .5 for all conditions of Experiment 1 (though it would have a slightly higher value for the spaced conditions of Experiment 3).

The effects of changing the number of categories, number of stimuli, or maximum stimulus frequency are predicted (Figures 7 and 8) without having to infer new values of w for each new number of categories or stimuli. This should be remembered when comparing these fits with those of the simple range-frequency model (Figures 5 and 6). Neither model does well with the two-category psychometric functions or the 100-point scales; the data for two categories (and also some features of the three-category scales) would be better fitted using the Thurstone-based, limen-centered model for range-frequency theory (Parducci, 1965); and the 100-point scales seem closer in form to magnitude estimations—except with simultaneous presentations. However, the fits are encouraging for all the other category scales. It is the ability of the elaborated model to predict the general magnitude of skewing effects and the general forms of the rating scales that encourages incorporation of the rule of consistency into the procedure for calculating frequency values.

Broader Implications

Although one might think that stimuli rarely repeat themselves exactly outside the laboratory (so that any skewing would be of spacing rather than of frequencies), the limits on discriminability along a single dimension (Miller, 1956) could mean that the effective frequencies would also vary. Thus one may expect to find the Category Effect in more general situations, for example, in polls of public opinion. Indeed, we have been finding it in our own (unpublished) research on ratings of the mental disturbance in clinical case histories, of the happiness expressed by schematic drawings of faces, and of the happiness of life events as expressed in short verbal descriptions.

Our attempts to account for Category and Stimulus Effects have forced us to consider how the effective context for judgment may be only a partial representation of the set we present in our experiments. A major attraction of employing psychophysical stimuli for the study of contextual effects has been that the experimenter seems to have so much control over the context. In past research on range-frequency theory, one could assume that the subject's context included the entire distribution of stimuli presented in the same experimental session. Now it appears more useful to assume that the effective context is based on only a small number of recent presentations and that as the number of categories increases, differences between stimulus frequencies are also reduced.

It is provocative to think of how the processes of retrieving past experiences from memory may influence judgment. Certainly, many of the situations to which one might want to apply range-frequency theory involve contexts established by earlier experiences (see Wedell & Parducci, 1985, for a discussion of possible applications). Although the specific restrictions imposed to explain the data from psychophysical experiments,

such as the limitation of working memory to the 12 most recent stimuli, may be tied to our particular experimental conditions, the notion that the effective context for judgment may be restricted to only a few stimulus events seems a more general possibility.

The principle of consistency would also seem to have a more general applicability. Whether a judgment describes the size of a square, the height of a person, or the aesthetic value of a poem, less information is transmitted when different judgments are applied to the same object.

The relational approach to judgment does assume a change in judgment when the same object is evaluated in a different context. The judgment describes the relation between the object that is judged and a context of similar objects. What the present research seems to show is that considerations of memory and consistency also determine which past experiences will constitute the effective context for judgment.

References

- Atneave, F. (1959). *Applications of communication theory to psychology*. New York: Holt, Rinehart & Winston.
- Bevan, W., & Avant, L. L. (1968). Response latency, response uncertainty, information transmitted and the number of available judgmental categories. *Journal of Experimental Psychology*, *76*, 394-397.
- Birnbaum, M. H. (1974). Using contextual effects to derive psychophysical scales. *Perception & Psychophysics*, *15*, 89-96.
- Braida, L. D., & Durlach, N. I. (1972). Intensity perception: II. Resolution in one interval paradigms. *Journal of the Acoustical Society of America*, *51*, 483-495.
- Garner, W. R. (1962). *Uncertainty and structure as psychological concepts*. New York: Wiley.
- Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.
- Hasher, L., & Zacks, R. (1979). Automatic and effortful processes in memory. *Journal of Experimental Psychology: General*, *108*, 356-388.
- Haubensak, G. (1982). An extension of Parducci's range-frequency theory of absolute judgments. In H.-G. Geissler & P. Petzold (Eds.), *Psychophysical judgment and the process of perception* (pp. 233-238). Berlin: VEB Deutscher Verlag der Wissenschaften.
- Helson, H. (1964). *Adaptation-level theory*. New York: Harper & Row.
- Mellers, B. A., & Birnbaum, M. H. (1982). Loci of contextual effects in judgment. *Journal of Experimental Psychology: Human Perception and Performance*, *8*, 582-601.
- Miller, G. A. (1956). The magical number seven plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*, 81-97.
- Parducci, A. (1956). Direction of shift in the judgment of single stimuli. *Journal of Experimental Psychology*, *51*, 169-178.
- Parducci, A. (1963). The range-frequency compromise in judgment. *Psychological Monographs*, *77*(2, Whole No. 565).
- Parducci, A. (1965). Category judgment: A range-frequency model. *Psychological Review*, *72*, 407-418.
- Parducci, A. (1982). Category ratings: Still more contextual effects. In B. Wegener (Ed.), *Social attitudes and psychophysical measurement* (pp. 262-282). Hillsdale, NJ: Erlbaum.
- Parducci, A. (1983). Category ratings and the relational character of judgment. In H.-G. Geissler & V. Sarris (Eds.), *Modern trends in perception* (pp. 89-105). Berlin: VEB Deutscher Verlag der Wissenschaften.
- Parducci, A., Calfee, R. C., Marshall, L. M., & Davidson, L. P. (1960). Context effects in judgment: Adaptation level as a function of the mean, midpoint, and median of the stimuli. *Journal of Experimental Psychology*, *60*, 65-77.
- Parducci, A., & Marshall, L. M. (1961a). Context effects in judgments of length. *American Journal of Psychology*, *74*, 576-583.
- Parducci, A., & Marshall, L. M. (1961b). Supplementary report: The effects of the mean, midpoint, and median upon adaptation level in judgment. *Journal of Experimental Psychology*, *61*, 261-262.
- Parducci, A., & Perrett, L. F. (1971). Category rating scales: Effects of relative spacing and frequency. *Journal of Experimental Psychology Monograph*, *89*, 427-452.
- Riskey, D. R., Parducci, A., Beauchamp, G. K. (1979). Effects of context in judgments of sweetness and pleasantness. *Perception & Psychophysics*, *26*, 171-176.
- Tabachnick, B., & Parducci, A. (1973). Improved recognition with feedback: Discriminability and range-frequency effects. *Bulletin of the Psychonomic Society*, *1*, 56-58.
- Thurstone, L. L. (1927). A law of comparative judgment. *Psychological Review*, *34*, 273-286.
- Tversky, A., & Kahneman, D. (1974). Judgment and uncertainty. *Science*, *185*, 1124-1130.
- Wedell, D. H. (1984). *A process model for psychophysical judgment*. Unpublished doctoral dissertation. University of California, Los Angeles.
- Wedell, D. H., & Parducci, A. (1985). Category and stimulus effects: A process model for contextual memory in judgment. In G. d'Ydewalle (Ed.), *Cognition, information processing, and motivation* (pp. 55-70). Amsterdam: North-Holland.

Received July 15, 1985

Revision received February 21, 1986 ■