Journal of Experimental Psychology
1971, Vol. 88, No. 2, 158–170

CONTEXTUAL EFFECTS IN INFORMATION INTEGRATION[1]†

MICHAEL H. BIRNBAUM,[2] ALLEN PARDUCCI, AND ROBERT K. GIFFORD
University of California, Los Angeles

Category judgments of the average lengths of sets of lines were inconsistent with context-independent models of information integration: the effects of any particular line upon the judgment of average length varied inversely with the lengths of the other lines within the same set. This interaction, obtained in five separate experiments, was similar to that previously reported for auditory intensities. The judgments reflect two kinds of contextual effects: (*a*) *within-set* effects, in which the judgment of the set varies directly with the range of values within the set, and (*b*) *between-set* effects, in which the apparent interaction between the stimuli within a set depends upon the context provided by the different sets. A simple range model provides a method for separating the two types of contextual effects. The context between sets is postulated to affect only the response scale; when the responses are rescaled to allow for the between-set context, the integrated impression is dependent upon both the mean and the range of components within the set.

The term "information integration" refers to the process whereby the psychological values of several stimuli are combined to produce a single impression. For example, *S*s have been asked to judge the overall loudness of four bursts of noise of varied intensity (Parducci, Thaler, & Anderson, 1968). Additive models of information integration assume that the integrated impression is simply a weighted sum (or average) of the values associated with each of the component stimuli comprising the set. In the usual application of such models, the effect of each component stimulus is assumed to be independent of the other stimuli. The term "additive models" will be used here to represent this assumption of independence from context.

In apparent contradiction to additive models, the loudness study demonstrated that the effect of the intensity at any one serial position within the set was greater when the intensities in the other three serial positions were lower. This finding was also inconsistent with the results of information-integration studies using verbal descriptions of personality (Anderson, 1968b). Consequently, the first aim of the present research was to determine whether these within-set effects are peculiar to judgments of loudness.

Experiment I: Replication with Visual Length

This experiment followed the design of the loudness study as closely as possible; however, the stimuli composing each set were lines of varying length presented simultaneously.

Method

Stimuli.—Sets of four black lines were centered 13 mm. apart and parallel to the longer side of 100 × 150 mm. white cards.[3] The length of the line in each position was either 25, 38, 51, or 64 mm., with a separate card for each of the 256 permutations of these four lengths.

Subjects.—Eight University of California, Los Angeles, undergraduates with no prior experience in this task were paid to serve as *S*s.

Procedure.—The *S*s were instructed to judge the average length of the four lines on each card, in comparison with the averages for the sets of lines on the other cards. Judgments were made on a 6-point scale from "1—Very Short to 6—Very Long." Each successive card was placed on the table before a

[1] This research was supported by United States Public Health Service Grant HD-00923, with the computing assistance of the University of California, Los Angeles Health Sciences Computing Facility, sponsored by National Institutes of Health Special Resources Grant RR-3. The authors wish to thank Norman H. Anderson, who made many helpful comments and suggestions on an earlier draft.

[2] Requests for reprints should be sent to Michael H. Birnbaum, Psychology Department, University of California, Los Angeles, California, 90024.

[3] Stimulus values for all of these experiments have been converted from the original English units and consequently are subject to rounding errors.

single *S* and removed as soon as he reported his judgment, usually within 5 sec. Each block of 64 sets in the regular series was as representative as possible of the total series. Order within blocks was random. A warm-up of 18 sets was representative of the total series with respect to physical averages, the different lengths appearing with equal frequency. Each *S* went through all 256 cards twice, once on each of 2 successive days. Individual sessions took a maximum of 40 min.

Results and Discussion

Each panel of Fig. 1 plots the judgments tabulated with respect to the lengths of the lines in two of the four positions. Position 1 refers to the top line on a card, Position 4 to the bottom line. Each point is a mean of 256 judgments: one for each of the 16 permutations of lengths for two of the positions, on each of 2 days, by each of eight *S*s. The upward trend of each curve represents the effects of the length of the line in the lower of the two positions; the separation between the curves represents the effects of the length of the line in the upper position.

Deviations from parallelism represent within-set contextual effects. These deviations are systematic and similar to the interactions obtained in the previous experiment on judgments of loudness. Again, the curves converge toward the right. The greater the length in one position, the less the effects of the length in the other position. This convergence characterizes the individual data (graphed as in Fig. 1) for each of the eight *S*s.

The significance of these interactions was tested by an overall analysis of variance in which the positions were the independent factors, each position having four levels (lengths) in a $4 \times 4 \times 4 \times 4$ design. The mean squares for each two-way interaction and its error term are given in each of the panels in Fig. 1. Since the critical value of F (9, 63) at the .001 level is approximately 3.8, the interactions between adjacent lines (Positions 1 and 2, 2 and 3, 3 and 4) are in each case significant. The other three two-way interactions are not significant ($p > .05$); of the five higher order interactions, only the $1 \times 2 \times 3$ interaction is statistically significant ($p < .01$).

The similarity between the results of line averaging and those of the previous experiment with auditory intensities strengthens the case against additive models. These models imply that the interactions in the analysis of variance are zero (Anderson,

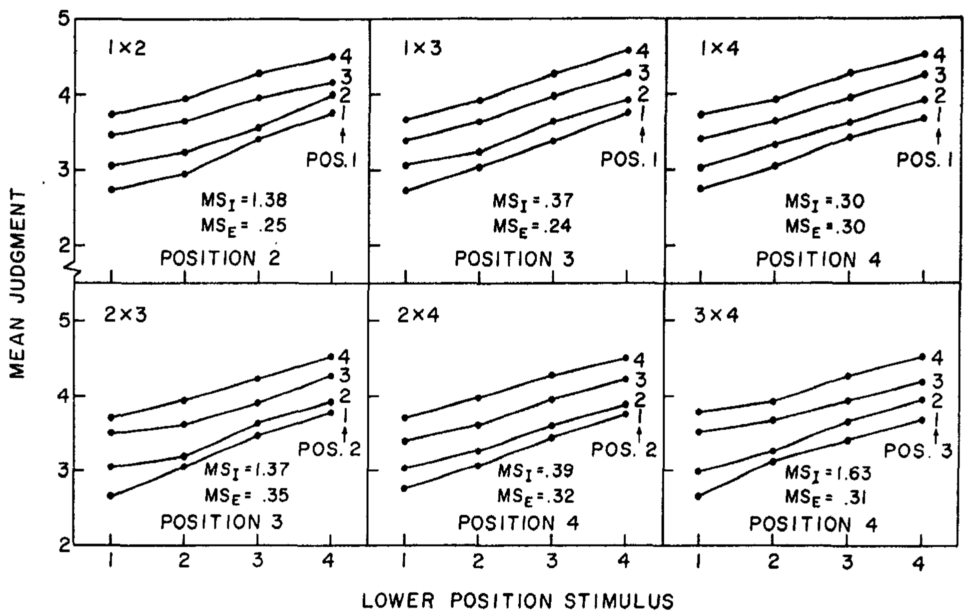

FIG. 1. Two-way data tables for each pair of positions. (Each data point is averaged over the 16 permutations of lengths at the other two positions.) (Exp. I)

1968b). Consequently, they are contradicted by the convergence found in the present experiment as well as in the previous experiment with auditory stimuli.

Experiment II: The Within-Set Context

Several alternative interpretations have been suggested to explain the within-set effects found for judgments of loudness (Parducci et al., 1968). One of these assumes that implicit judgments of the components of each set are determined by the same contextual conditions as overt judgments of single stimuli. In particular, the implicit judgment of each component of a set is assumed to vary inversely with the lengths of the other components in the set. This is the *contrast* effect typically found in judgments of single stimuli. If the overt judgment of the set were simply a weighted average of these implicit judgments, convergence interactions would be obtained, particularly if contrast were greater for longer lines.

To test this interpretation directly, the present experiment required some *S*s to make separate judgments of each of the components of the sets while other *S*s judged the average length of the set, as in Exp. I. The component judgments should reflect the postulated contrast effects directly, and the mean of the judgments of the components of a set should predict the judgment of the average length of the set.

The number of components was reduced from four to two. This permitted an increase in the number of lengths for each position from four to eight without making the series too long for replication within a single experimental session. Furthermore, the use of only two lines in a set eliminates a number of complex interpretations, such as the possibility that *S* ignores shorter lines in Positions 2 and 3 when longer lines are presented in Positions 1 and 4.

Method

Stimuli.—Sets of two black lines were centered 13 mm. apart and parallel to the longer sides of 130

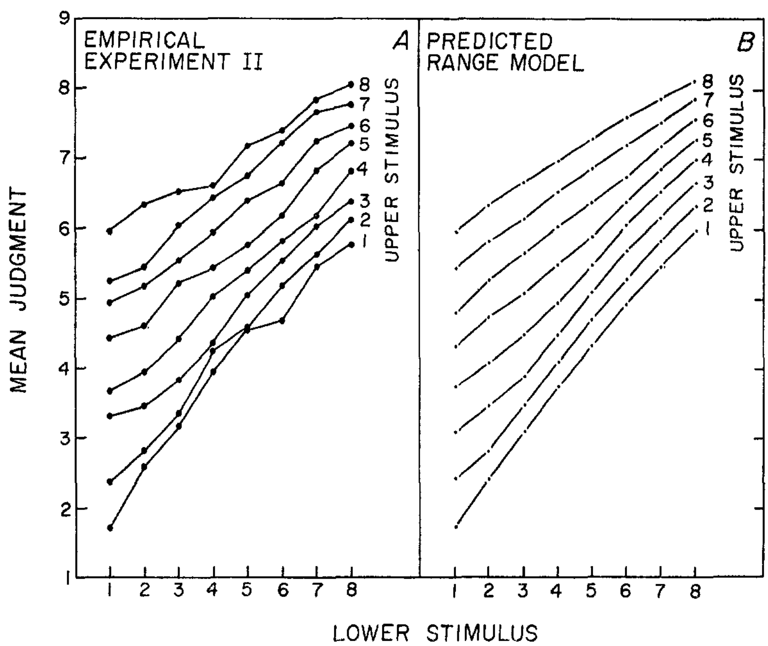

FIG. 2. (*a*) Mean judgment of average length, plotted as joint function of lengths of lower and upper lines in the set; additive models predict parallelism for these curves, and (*b*) predictions from the range model described in the Discussion section, for the same conditions. (Exp. II)

× 200 mm. white cards. The eight lengths varied from 25 to 70 mm., in 6-mm. steps, with a separate card for each of the 64 permutations. The cards were exposed manually for approximately 5 sec. each, at a distance of from 2 to 4 m.

Subjects.—Eighty University of California, Los Angeles, undergraduates were run for single sessions in eight groups of from 8 to 15 *S*s each. As in each of the following experiments, *S*s were fulfilling a requirement of the introductory course in psychology. No *S* served in more than one experiment or in more than one experimental condition.

Procedure.—Four of the groups, the 44 *S*s judging average length, were instructed to compare the average length of the two lines on each card with the averages of the lengths on the other cards, as in Exp. I. These *S*s made a single overall judgment for each card. The rest, the 36 *S*s judging the individual components, were instructed to "judge the length of each line compared to all the other lines, not just relative to the other line on the same card"; they were thus required to make two separate judgments for each card, judging the upper line first, then the lower line. In both conditions, *S*s recorded numerical judgments for categories running from "1—Very Very Short" to "9—Very Very Long." After a warm-up of 18 representative presentations, the 64 cards were presented in random order and then again in reverse order. A different order was used for each of the four groups in each condition.

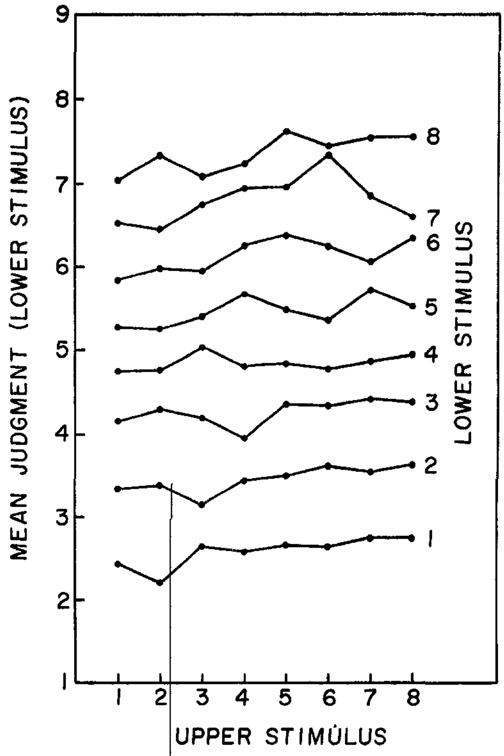

FIG. 3. Mean judgment of lower line as function of length of upper line in same set. (Exp. II)

Results and Discussion

Figure 2*a* plots the mean judgments of the "average length" for each set, tabulated as in Fig. 1. Again, there is a significant convergence to the right, F (49, 2107) = 8.02, $p < .001$. Most of the variance of the interaction (71%) is in the Linear × Linear trend. As in Exp. I, additive and averaging models predict no interactions, i.e., parallel curves. This prediction is again contradicted by the convergence in Fig. 2*a*. Figure 2*b* represents theoretical predictions from a range model that will be developed in the final discussion section.

Figure 3 plots the separate judgment of the lower line as a function of the length of the upper line on the same card. Contrary to hypothesis, the linear trend of these functions is positive rather than negative; judgments of the individual lines vary *directly* rather than *inversely* with the length of the other line on the same card, F (1, 35) = 39.42, $p < .001$. Figure 3 suggests that this overall assimilation is complicated by a contrast for pairs of similar lengths. Thus, the most prominent dips in these functions occur where the length of the lower stimulus is one step below the length of the upper stimulus. The interaction is statistically significant, F (49, 1715) = 3.15, $p < .001$, with much of the variance in the quadratic trends. Highly similar effects were found for judgments of the upper lines.

Figure 4 plots the mean of the two separate judgments for each pair. Although the interaction is statistically significant, F (49, 2445) = 1.97, $p < .001$, it is small and without the marked convergence of Fig. 2*a*. Convergence would have been obtained if the component judgments, as represented in Fig. 3, had shown both contrast and convergence to the right, i.e., a contrast that was greater for the longer lines. Neither the separate judgments nor their mean seem consistent with the contrast interpretation. The difference between the curves shown in Fig. 2*a* and 4 is

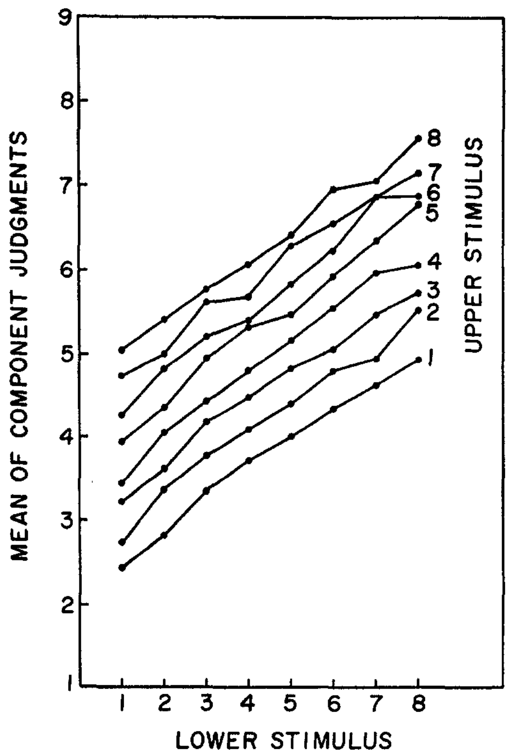

FIG. 4. Mean of the separate judgments of component lengths, plotted for each permutation as in Fig. 2. (Exp. II)

inconsistent with the hypothesis that the average of component ratings would explain the rating of the average.

It is possible that the request for component judgments reduces the importance of the other line in the same set. Another possibility, consistent with previous data (Anderson, 1966; Anderson & Lampel, 1965; Wyer & Dermer, 1968), is that overt judgments of the components reflect a confusion by *S* between the component he is supposed to be judging and his overall impression of the set. Although either of these possibilities might have obscured a real within-set contrast operating on implicit judgments, the fact that differences in component ratings are generally in the direction of assimilation rather than contrast encourages consideration of alternative interpretations.

Figure 5 plots the judgments of the sets in Exp. II as a function of within-set range, i.e., the difference between the lengths of the two components of each set. The physical lengths average to the same value for each point on the abscissa so that the range and mean of the lengths are not confounded in Fig. 5. The linear component of this function is statistically significant, $F(1, 43) = 28.62$, $p < .001$, judgments varying directly with range. Similar results were obtained for judgments of the four-component sets of Exp. I. This suggests that the integrated impression depends upon the within-set range.

Experiment III: The Within-Set Range Effect in Comparative Judgment

In this experiment, the hypothesis of a within-set range effect is tested directly by having *S*s compare the average length of a pair of lines with the length of a single line.

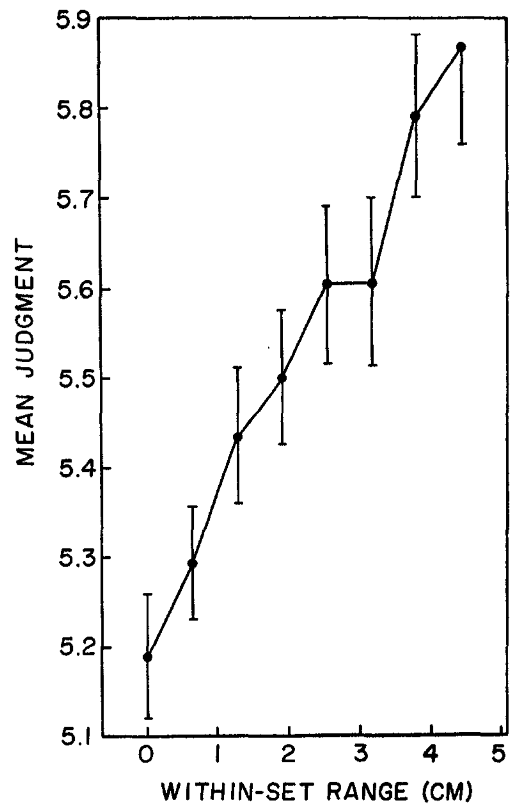

FIG. 5. Mean of judgments of those sets having the same range of lengths. (Each vertical extends ± 1 *SE*.) (Exp. II)

The psychological average of each pair is then the length of the single line to which it is equated. The experimental design varies the within-set range for different pairs of lines while controlling for mean physical length.

Method

Stimuli.—Nine standards and 12 comparison stimuli were used. The standards were single lines drawn lengthwise and centered on 130 × 200 mm. white cards; their lengths varied from 41 to 54 mm., in 1.6-mm. steps. The comparison pairs were chosen from the sets used in Exp. II, with each pair again on a separate card. For each of three mean lengths (44, 48, and 51 mm.), there were four comparison pairs differing in within-set range: (44, 44), (38, 50), (31, 57), (25, 63); (45, 51), (38, 58), (32, 64), (26, 70); (51, 51), (45, 57), (38, 64), (32, 70).

Subjects.—All 108 permutations of the standard and comparison stimuli were judged by each of 18 *S*s.

Procedure.—The task on each trial was to compare the average length of the two lines on the comparison card with the length of the single line on the standard card. The comparison categories on the 6-point scale were: 1. Shorter; 2. Slightly Shorter; 3. Very Slightly Shorter; 4. Very Slightly Longer; 5. Slightly Longer; and 6. Longer. Each standard remained on a stand 2 to 4 m. from *S*s, while the 12 comparison sets were placed next to it, 1 at a time, for approximately 5 sec. each. Different random orders of standards and comparisons were used for each squad of nine *S*s.

Results

Figure 6 plots the mean judgments of each of the 12 comparison pairs as a function of their difference in length (within-set range), but averaged over the nine standards. Each curve represents the judgments of 4 comparison pairs with the same physical mean. It is clear that these comparative judgments increase directly with difference in length, even when mean physical length is held constant. The linear trend of these functions is significant, $F(1, 17) = 6.64$, $p < .05$, which provides direct evidence for the within-set range effect inferred from the data of Exp. I and II.

The direction of the within-set range effect appears inconsistent with the averaging process described by the basic adaptation-level equations (Helson, 1964). These equate adaptation level or the psychological average to the geometric mean of the stimulus values. But since geometric mean varies inversely with range when arithmetic mean is held constant, the adaptation-level approach incorrectly implies that the functions in Fig. 6 should have negative slope rather than positive slope.

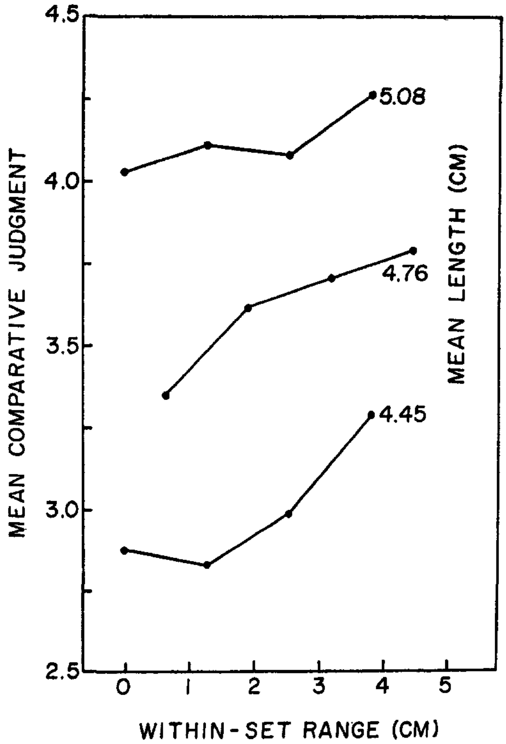

FIG. 6. Mean judgment of comparison set as joint function of range of lengths within set and mean length of components. (Exp. III)

EXPERIMENT IV: TEST OF THE WITHIN-SET RANGE EFFECT WITH CATEGORY RATINGS

The comparative judgments of Exp. III can also be analyzed with respect to the traditional psychophysical measure of central tendency, the point of subjective equality (PSE). This is the value of the standard to which each comparison set would have been equated (with the interpolated standard value corresponding to 3.5 on the 6-point scale of comparative judgment). The following equation describes the relationship of PSE to the physical values and range of the comparison set:

$$PSE = .5(\Phi_L + \Phi_S) + .25(\Phi_L - \Phi_S), \quad [1]$$

where Φ_L and Φ_S are the lengths of the longer and shorter of the two lines in the comparison set, and .25 is an empirically fitted constant reflecting the importance of the within-set range effect.

Insofar as Equation 1 provides a useful characterization of the contextual effects within sets, then category ratings should be the same for all line pairs having the same PSE. In particular, for each pair of lines with unequal lengths, there is a pair of lines of equal length which should receive the same judgment. To test this implication, two different series of stimuli are employed in the present experiment: (*a*) a 5 × 5 *factorial series* and (*b*) an *equal-component series* in which the length of each pair corresponds to the PSE for one of the sets in the factorial series. Insofar as the within-set effects are accounted for by the PSE equation from Exp. III, both series should have the same context between sets. Consequently, the judgments should be the same for corresponding sets in the two series.

Method

Stimuli.—The factorial series was composed of pairs of lines, arranged as in Exp. II but with each line taking only five values (in 10-mm. steps from 25 to 65 mm.). In the equal-component series, both lines in the pair had the same length. This length was the value of PSE from Equation 1 for the corresponding set in the factorial series.

Subjects.—Four groups (29 *S*s) judged the factorial series first, then the equal-component series; the remaining three groups (19 *S*s) judged the equal component series first, then the factorial series.

Procedure.—Each series was preceded by a warm-up of 12 presentations. After the 25 sets in each series had been presented in random order and again in reverse order, *S*s were told,

> Now there is another series of cards. Your task remains the same, i.e., to judge the average length of each set compared with the average length of the other sets; but now the two lines on each card will always be of equal length (or "may have different lengths").

The first replicate of each stimulus type was treated as additional warm-up, the regular warm-up being insufficient to prevent strong ordinal effects for some of the stimuli presented near the beginning of the series. Since tests for transfer between the two series proved insignificant, a within-*S*s design was employed in analyzing the results.

Results and Discussion

Comparison of Fig. 7a and 7b shows obvious similarity of form, verified by the nonsignificance of the Series × Upper × Lower interaction, F (16, 576) = 1.50, $p > .10$. Deviations from the best fit for the additive models can be represented by the sum of squares associated with the Upper × Lower interaction, which is almost six times as great as the sum of squares for the Series × Upper × Lower interaction. The nonsignificance of the other sources of variance associated with series is consistent with the prediction that the equal-component sets would receive the same judgments as the corresponding factorial sets. The results thus support the interpretation that apart from between-set effects, which were presumably the same for both series, interactions are determined by the within-set range as described by Equation 1.

Experiment V: The Context between Sets

In addition to the contextual effects associated with the range of stimuli within sets, interactions may also depend upon the context between sets. Previous research has shown that sets are assigned higher categories when the frequency distribution of their physical means is positively skewed, i.e., when the sets with lower physical means are presented more frequently (Parducci et al., 1968). In this respect, the judgments of sets are like those of single stimuli. A general finding with single stimuli is that when the frequency distribution of the stimuli is negatively skewed, the judgment function is positively accelerated; when the frequency distribution is positively skewed, the function is negatively accelerated (Parducci, 1965).

This suggests that experimental manipulation of the frequency distribution of sets would affect the *form* of the plot of the judgments of the sets against their physical

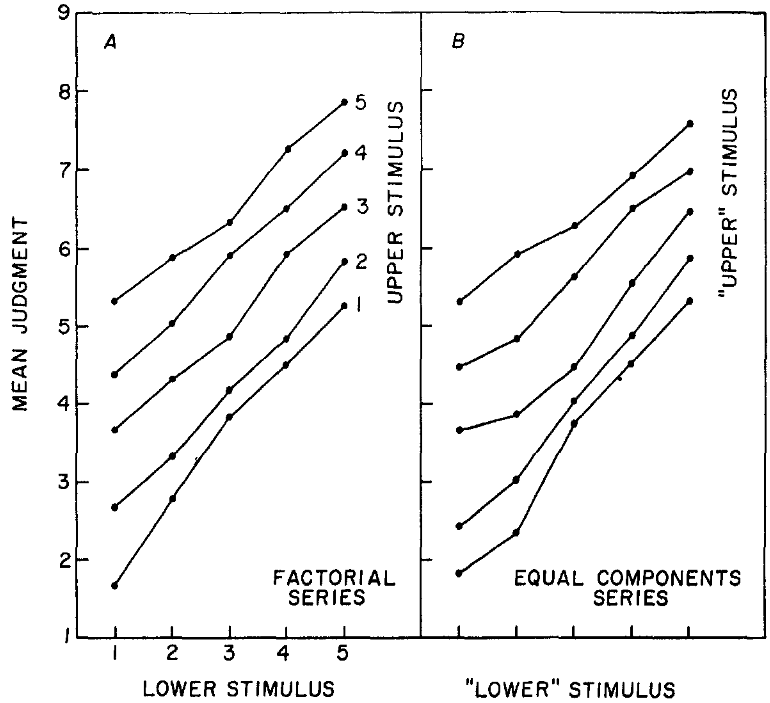

FIG. 7. Mean judgment of average length of each set. Coordinates for equal-component sets represent upper and lower lines for the factorial set whose PSE (calculated from Equation 1) corresponds to the lengths of the equal-component set. (Exp. IV)

means. Systematic nonlinearity in this plot implies interactions in the analysis of variance of the responses, assuming homogeneity of response to sets with the same physical mean. For example, if the judgment of a set with Physical Values 1 and 5 were closer to the judgment of a set both of whose physical values were 5 than to a set whose values were both 1, the judgments would show the convergent interaction obtained in Exp. I and II.

Apart from within-set contextual effects, any nonlinear change in the function relating the judgment of the set to the physical mean of the stimuli in the set must be paralleled by a change in this interaction. This would ordinarily be interpreted as evidence against the additive models. The change would be described as a "between-set effect" since it would be produced by manipulation of the frequency distribution of sets.

Experiment V manipulates the context by adding extra sets to those whose judgments are analyzed for interactions. Filler sets have been used for various purposes (e.g., Anderson & Jacobson, 1965), including an effort to break up whatever pattern may be present in the experimental sets. It is possible that the use of filler sets may affect or even obscure the interactions that might otherwise be obtained in research on information integration.

Method

Stimuli.—The sets from the 5 × 5 factorial series of Exp. IV were used as the regular stimuli. In addition to these 25 sets, 25 filler pairs were added for each of two contextual conditions. The means of the lengths of the filler pairs in the Low- and High-Filler conditions were approximately 38 and 52 mm., respectively. In the Low-Filler condition, the lengths in each filler pair (inside parentheses) and their frequencies of presentation were as follows: (35, 35)3, (25, 55)6, (35, 45)6, and (25, 45)10. In the High-Filler condition, these filler values were: (55, 55)3, (35, 65)6, (45, 55)6, and (45, 65)10.

Subjects.—The 30 *S*s in the High-Filler condition and the 22 in the Low-Filler condition were run in groups of from 8 to 11 *S*s.

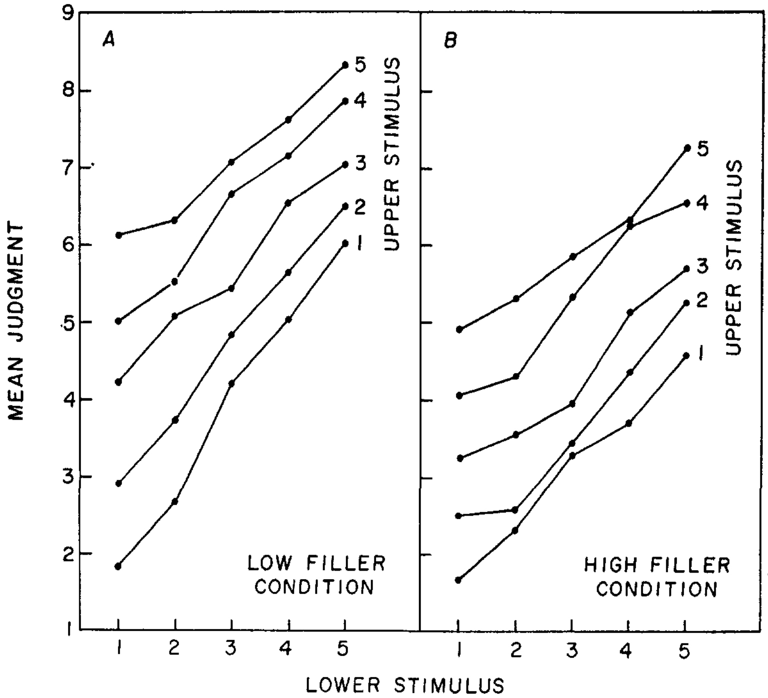

FIG. 8. Mean judgment of average length for Low-Filler and High-Filler contexts. (Exp. V)

Procedure.—Instructions and method of presentation followed the procedure used for the condition of Exp. II that required judgments of the average length of each set. The entire series was presented in random sequence, and then in reverse, with a different sequence for each group.

Results and Discussion

The mean judgments of the 25 sets common to both conditions are shown in Fig. 8. Convergence is much less for the High-Filler condition (Fig. 8*b*), than for the Low-Filler condition (Fig. 8*a*), and the Filler Context × Length of Upper Line × Length of Lower Line interaction is statistically significant, $F\ (16, 672) = 2.37$, $p < .01$. Although the Upper × Lower interaction is also significant for the High-Filler data analyzed separately, $F\ (16, 464) = 2.92$, $p < .001$, the shape of the interaction is clearly dependent upon the context established by the filler sets.

The two functions in Fig. 9 have different forms, and the Stimulus Mean × Filler Context interaction is highly significant, $F\ (8, 336) = 5.79$, $p < .001$, with about half of the variance in the quadratic trend. The Low-Filler condition has a negatively accelerated function with a highly significant quadratic trend, $F\ (1, 21) = 31.83$, $p < .001$, whereas the High-Filler function exhibits a slight but statistically insignificant positive acceleration. It is this nonlinearity that implies the interactions shown in Fig. 8.

When the data are segregated by within-set range (as for Fig. 5), the range variable is again significant, $F\ (4, 42) = 18.52$, $p < .001$. Its interaction with filler context is not significant, $F\ (4, 42) = 1.56$, $p > .10$. There is thus no evidence that the within-set range effect itself depends upon the context between sets. This suggests that the between-set context has an independent influence upon the interactions.

The major finding of Exp. V is that the statistical interactions ordinarily interpreted as within-set contextual effects are also determined by the context between sets. Although only the filler sets were manipulated in this experiment, it seems likely that the interactions in Fig. 8 were also influenced by the particular values selected for the regular factorial series. Thus, the context between sets must be taken into account in any interpretation of either the presence or absence of this type of interaction.

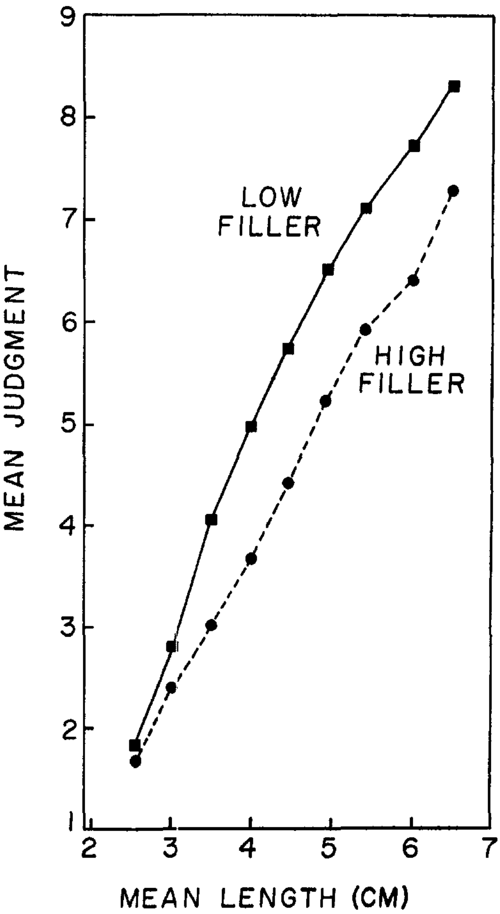

FIG. 9. Mean judgment of average length as function of mean length of components. (Exp. V)

Discussion

Figure 10 provides a general framework for discussing different interpretations of the data. In this framework, the physical values of the component stimuli, Φ_k, are transformed to psychological values, s_k, by the psychophysical function, H. These psychological values are combined by the integration rule, I, to form the overall impression, $\Psi_{1 \ldots k \ldots n}$, which is then transformed into the overt response, $R_{1 \ldots k \ldots n}$, by the judgment function, J.

Additive models.—Additive models of information integration assume that the psychological values are added or averaged to form an integrated impression. The best supported of these models (Anderson, 1968b) asserts the following function for I:

$$\Psi_{1 \ldots k \ldots n} = \sum_{k=1}^{n} w_k s_k, \qquad [2]$$

in which w_k is the weight of whatever stimulus is in Position k, and s_k is the psychological value of the stimulus in that position. In the usual application, it is assumed that the weights are independent of the stimulus values in the set, and the effects of each stimulus are assumed to be independent of the other stimuli with which it is combined. Although it is the overt response, R, rather than the psychological impression, Ψ, that is actually observed in experiments on information integration, the relationship, J, between these two variables is usually assumed to be linear. These assumptions permit the use of analysis of variance on

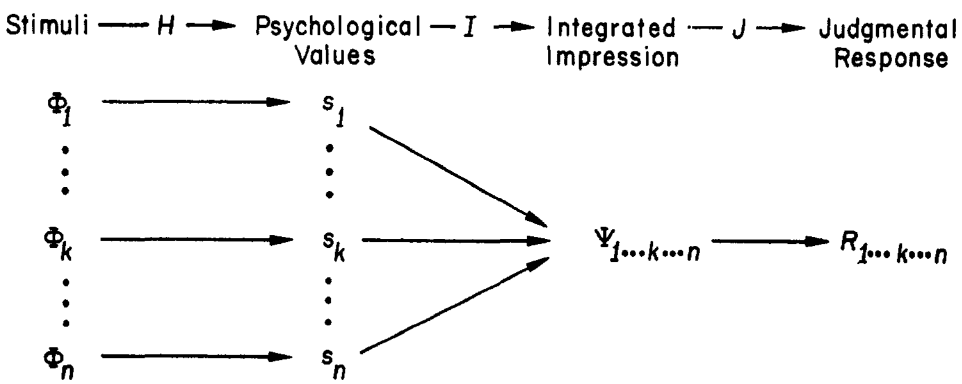

FIG. 10. Framework outline of information integration analysis.

the overt responses to test the additive models (Anderson, 1968b).

The convergence interaction obtained in each of the present experiments constitutes evidence against this class of additive models. The simplest interpretation is that the integration of information, at least in the present situations, is not simply an additive or averaging process. The effect of any line in the set on the overall impression depends on the other lines in the same set.

A simple additive or averaging rule might be preserved for I at the expense of some complication of other features of Fig. 10. It is unnecessary, for example, to assume that J is a linear function. Indeed, if it were the raw physical values that were averaged (or if H were a linear function), J might be expected to take the logarithmic form that describes the typical empirical relationship between category judgments and physical magnitudes of single stimuli (Stevens & Galanter, 1957). With the actual integration still conceived as an additive process, an antilog transform of the responses could account for part of the convergence interactions (Parducci et al., 1968). However, this possibility is contradicted by the finding (Exp. I and II) that judgments of sets with the same physical average vary directly with the within-set range. The within-set effect is also found when the task requires direct comparison of each set with a standard (Exp. III).

The possibility of a nonlinear function relating the integrated impression to the overt response need not be tied to physical averaging. If it were assumed that the psychological values, s_k, were averaged in accordance with the weighted-averaging model (see Equation 2), the elimination of interactions would be the appropriate criterion for evaluating alternative candidates for J (Anderson, 1970). However, the demonstration (Exp. V) that changing the filler sets also changes the interaction indicates that J would itself be subject to contextual effects. Even if there were no filler sets, J might be expected to depend on the other sets in the series presented for judgment, just as the form of the judgment function for single stimuli depends on the other stimuli in the series (Parducci, 1965).

Contextual effects might also precede the integration process. Thus, the psychological scale value of a component stimulus, s_k, might depend not just on its own physical value, Φ_k, but also on the other components, $\Phi_1, \Phi_2, \ldots, \Phi_n$, of the same set. The integrated impression might then be a weighted average of the contextually determined scale values. However, the judgments of the components in Exp. II did not average to the overall judgment of the set.

Another possibility, one that has seemed attractive for interpreting certain interactions with stimuli that are more complex (e.g., Lampel & Anderson, 1968), attributes the interaction to changes in the weights rather than to changes in the scale values: the value of w_k in Equation 2 might depend upon the value of s_k. For example, the weight or relative influence of a particular line in the set might be directly proportional to its length. Preliminary attempts to fit the present data by permitting the weights to vary in this manner have not been encouraging.

A range model.—The data from these experiments can be described more simply by assuming a nonadditive integration of physical values. The integration function, I, is expressed in terms of the physical values, bypassing the need for H; and the integrated impression, Ψ, is assumed to vary directly with the range of the set. Thus, the within-set range effect is ascribed to the integration process.

The effect of the context between sets is attributed by this model to the final transformation, J. Once Ψ is achieved for a particular set, it is compared with the values of Ψ for the other sets in the series. These different Ψ values constitute a frequency distribution in which sets with greater within-set ranges have relatively higher values. The judgment assigned to each value in this distribution presumably follows the principles describing judgments of single stimuli. The general trend of the differences between the Low- and High-Filler conditions of Exp. V are consistent with the contextual effects found with single stimuli (Parducci, 1965), as were the between-set effects in the experiment with auditory stimuli.

Equation 1, which was fitted to the comparative judgments of Exp. III, was a preliminary candidate for the integration function, I. The cross-validation of Equation 1 in Exp. IV demonstrated that the within-set range could account for the interaction. However, the much more extensive explorations of the other experiments suggest a modification of Equation 1. In particular, a post hoc analysis of Exp. II (by connecting all points in Fig. 2*a* representing judgments of sets with the same physical mean) suggests that the effect of within-set range is inversely proportional to the mean of the components.

Accordingly, Equation 1 can be modified as follows to provide a better account of the within-set range effects:

$$\Psi = .5(\Phi_L + \Phi_S) + \omega(\Phi_L - \Phi_S)/(\Phi_L + \Phi_S), \quad [3]$$

in which Ψ is the psychological average of the longer and shorter of the component lengths, Φ_L and Φ_S respectively, and ω is an empirical constant representing the magnitude of the within-set range effect.

The value of ω and also the relationship between Ψ and the actual judgment are estimated from empirical data. First, the judgments of those pairs with equal components (and thus no within-set range) are fitted using the least-squares criterion for the cubic function:

$$R = \alpha\Psi + \beta\Psi^2 + \lambda\Psi^3 + \delta. \quad [4]$$

The inverse function for Equation 4 is then applied to transform the judgment of each set into its corresponding Ψ value. The purpose of this is to allow for between-set effects. These Ψ values and the physical lengths for each pair are then substituted in Equation 3 to obtain the best least-squares estimate of ω. This fit to the data requires the estimation of only the four parameters of Equation 4, in addition to ω. For the data of Fig. 2*a*, these five parameters (with $\omega = .823$) yield the predictions shown in Fig. 2*b*. The correspondence between Fig. 2*a* and 2*b* represents a marked improvement over the fit obtained for the additive models, even when the judgments are first rescaled using Equation 4 and the usual analysis of variance procedures are employed to fit these same data (i.e., with parallel lines). The variance left unexplained by the additive models is closely associated with the range term from Equation 3. The range model does significantly better than additive models, $F(1, 61) = 60.37$, $p < .001$, reducing the unexplained variance to less than one-third of the variance left unexplained by the additive models. An even greater improvement of fit is obtained for the Low-Filler condition of Exp. V. Additive models do a better job only on the High-Filler condition where their fit is very good. However, when ω is estimated separately for each physical mean, the range model reduces the error variance of the High-Filler condition to one-half of that from the additive models, still using fewer parameters. It should be emphasized that the effect of between-set context in the range model is estimated solely from sets with equal components.

Equation 4 is assumed to be the function, J, that transforms the psychological impressions to the responses. Application of J^{-1} can be considered a method for rescaling the responses. The fact that significant interactions remained after this rescaling suggests that the within-set range effects are "real" and not an artifact of the response scale.

Concluding comments.—There are numerous differences in procedure between the present experiments and those that are more consistent with the weighted-averaging or additive models. For example, in order to isolate the integration process from the problems of shifting scales of judgment, the between-set context has sometimes been anchored by end-sets that are more extreme than the sets whose judgments are analyzed in tests of these models. Before judging the regular series, *S* may be shown two extreme sets and told they represent "1" and "20" on a 20-point scale. There is some evidence that such anchoring reduces the interactions (Anderson, 1967b). But insofar as variability of judgment for each stimulus is proportional to the range of all the stimuli, the power of the test of the interactions is also reduced.

A recent experiment (Weiss & Anderson, 1969) used the method of reproduction to study averaging of serially presented lines. The interactions were described as slight. This method may reduce the effects of the between-set context, but in preliminary work[4] we have obtained large convergent interactions using the method of reproduction and stimuli similar to those of Exp. II. There were other differences in procedure, including the use of only two alternative lengths in each position in the Weiss and Anderson study. Earlier psychophysical support for the weighted-averaging model used lifted weights, just two different values for the component stimuli, and a 20-point rating scale (Anderson, 1967a). Although it would be difficult to pin down the effects of the procedural differences, the present experiments give substantial evidence that the averaging and additive models may hold only under very restricted conditions.

Interactions reflecting either the between-set or within-set contexts do not appear to be restricted to judgments of lines and noise bursts. They are prominent in our unpub-

[4] Birnbaum, M. Studies in information integration, unpublished manuscript, 1967.

lished work on lifted weight, (see Footnote 4) on the favorableness of adjectives, and on the morality of acts of behavior. Evidence consistent with the within-set range effect is also found in published studies of information integration (Anderson, 1965, 1968a; Lampel & Anderson, 1968; Weiss, 1963; Willis, 1960). Although some of the interactions reported in these studies have been interpreted as shifts in the weighting of different classes of stimuli, they may also be interpreted as within-set range effects.

The present research demonstrates that there are conditions in which one must expect strong contextual effects upon information integration. The effect of a particular component stimulus upon the judgment of a set depends upon the range of component values. The judgment of the set also reflects the context established by the differences between sets.

REFERENCES

ANDERSON, N. H. Averaging versus adding as a stimulus-combination rule in impression formation. *Journal of Experimental Psychology*, 1965, **70**, 394–400.

ANDERSON, N. H. Component ratings in impression formation. *Psychonomic Science*, 1966, **6**, 279–280.

ANDERSON, N. H. Application of a weighted average model to a psychophysical averaging task. *Psychonomic Science*, 1967, **8**, 227–228. (a)

ANDERSON, N. H. Averaging model analysis of set-size effect in impression formation. *Journal of Experimental Psychology*, 1967, **75**, 158–165. (b)

ANDERSON, N. H. Application of a linear-serial model to a personality-impression task using serial presentation. *Journal of Personality and Social Psychology*, 1968, **10**, 352–362. (a)

ANDERSON, N. H. A simple model for information integration. In R. P. Abelson et al., *Theories of cognitive consistency: A source book.* Chicago: Rand McNally 1968. (b)

ANDERSON, N. H. Functional measurement and the psychophysical law. *Psychological Review*, 1970, **77**, 153–170.

ANDERSON, N. H., & JACOBSON, A. Effect of stimulus inconsistency and discounting instructions in personality impression formation. *Journal of Personality and Social Psychology*, 1965, **2**, 531–539.

ANDERSON, N. H., & LAMPEL, A. K. Effect of context on ratings of personality traits. *Psychonomic Science*, 1965, **3**, 433–434.

HELSON, H. *Adaptation-level theory.* New York: Harper & Row, 1964.

LAMPEL, A. K., & ANDERSON, N. H. Combining visual and verbal information in an impression-formation task. *Journal of Personality and Social Psychology*, 1968, **9**, 1–6.

PARDUCCI, A. Category judgment: A range-frequency model. *Psychological Review*, 1965, **72**, 407–418.

PARDUCCI, A., THALER, H., & ANDERSON, N. H. Stimulus averaging and the context for judgment. *Perception & Psychophysics*, 1968, **3**, 145–150.

STEVENS, S. S., & Galanter, E. H. Ratio scales and category scales for a dozen perceptual continua. *Journal of Experimental Psychology*, 1957, **54**, 377–411.

WEISS, D. J., & ANDERSON, N. H. Subjective averaging of length with serial presentation. *Journal of Experimental Psychology*, 1969, **82**, 52–63.

WEISS, W. Scale judgments of triplets of opinion statements. *Journal of Abnormal and Social Psychology*, 1963, **66**, 471–479.

WILLIS, R. H. Stimulus pooling and social preception. *Journal of Abnormal and Social Psychology*, 1960, **60**, 365–373.

WYER, R. S., & DERMER, M. Effect of context and instructional set upon evaluations of personality-trait adjectives. *Journal of Personality and Social Psychology*, 1968, **9**, 7–14.

(Received June 15, 1970)