JOURNAL OF EXPERIMENTAL SOCIAL PSYCHOLOGY 23, 230–249 (1987)

A Formal Analysis of Ratings of Physical Attractiveness: Successive Contrast and Simultaneous Assimilation

DOUGLAS H. WEDELL, ALLEN PARDUCCI, AND R. EDWARD GEISELMAN

University of California, Los Angeles

Received June 27, 1986

Photographs of faces were presented in a series, either singly or in pairs, for ratings of physical attractiveness. In Experiment 1, faces were presented singly, and both the range and relative frequencies of physical attractiveness (on baseline scaling) were manipulated experimentally. The same face elicited higher ratings when less attractive faces predominated in the experimental series, *successive contrast*. Increasing the number of available categories resulted in higher ratings but did not reduce the amount of successive contrast. Both range and skewing effects were in accordance with a range-frequency model that permits the subjective range to vary with number of categories. In Experiment 2, faces were presented in pairs. The same face now elicited lower ratings when presented simultaneously with a less attractive face, *simultaneous assimilation*. Successive contrast was again observed between pairs and was greater for 5- than for 101-point rating scales. A model that uses the judgments resulting from a range-frequency compromise as the stimulus values for integration within pairs provides the best account of how both contrast and assimilation occur within the same experimental session. Alternative interpretations of the observed contrast and assimilation were discussed. © 1987 Academic Press, Inc.

The same face is judged more attractive when presented in a series with less attractive faces than when presented with more attractive faces (Kenrick & Gutierres, 1980; Kernis & Wheeler, 1981; Melemed & Moss, 1975). This has been called "contrast," and it is generally consistent with contextual theories of judgment (e.g., Beebe-Center, 1932; Helson, 1964; Parducci, 1983). However, "assimilation" has also been reported: The same face is judged to be more attractive when presented alongside attractive faces than when presented with unattractive faces (Geiselman, Haight, & Kimata, 1984; Melemed & Moss, 1975; Strane & Watts, 1977).

It seems paradoxical to find *both* contrast and assimilation in ratings of physical attractiveness. This paper attempts to resolve this paradox

Requests for reprints should be sent to Dr. Douglas Wedell, Department of Psychology, University of Illinois, 603 E. Daniel Street, Champaign, IL 61820.

0022-1031/87 $3.00

by showing that the direction of the contextual effect depends on how stimuli are presented, different stimulus arrangements determining different contextual parameters. Contrast is interpreted as the result of comparison processes that locate the stimulus within the distribution of successively presented stimuli. Assimilation is interpreted as a failure to separate the individual stimulus from other stimuli that are simultaneously present. Experiments producing assimilation have generally presented contextual stimuli simultaneously with the target stimulus. This type of presentation results in the averaging of all stimuli simultaneously present.

To study contrast effects, Experiment 1 manipulates two aspects of the contextual distribution (range and relative frequency of stimuli). These effects are evaluated in terms of Parducci's (1983) range-frequency theory of judgment. In Experiment 2, the immediate context is manipulated to produce assimilation, while the overall context is manipulated to produce contrast. These effects are evaluated in terms of models that combine range-frequency theory with Anderson's (1981) integration theory.

EXPERIMENT 1: RANGE AND FREQUENCY MANIPULATIONS FOR SUCCESSIVE STIMULI

Contrast may be understood as an effect upon judgment of the distribution of contextual stimuli. Parducci's (1963, 1965, 1983) range-frequency theory asserts that the judgment of a stimulus reflects its place in the distribution of contextual stimuli in accordance with a compromise between two principles of judgment, both of which can contribute to contrast. According to the range principle, subjects divide the stimulus range into subjectively equal intervals corresponding to the number of categories in their rating scale. Since the rating of any stimulus is determined by the proportion of the range below it, the range principle can account for contrast effects that occur when the stimulus range differs between contextual sets.

According to the frequency principle, subjects assign the same number of stimuli to each rating category. This means that the rating of any stimulus is determined by the proportion of contextual stimuli below it. The frequency principle accounts for contrast effects that occur when the relative frequencies of stimuli differ between contextual sets with the same stimulus endpoints.

Range-frequency theory represents the judgment of a stimulus by a weighted average of range and frequency values:

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

where C_{ic} is the overt category rating of Stimulus i in Context c, R_{ic} is its rating implied by the range principle, F_{ic} is its rating implied by the frequency principle and w is the weighting factor. For a variety of psychophysical dimensions (e.g., size, numerical magnitude, sweetness of taste), good fits have been achieved with w close to 0.5, representing

TABLE 1
Design of Experiment 1: Frequencies for Each Stimulus Group

		Restricted range		Full range	
Group	Range[a]	Positive	Negative	Positive	Negative
A	1.20–1.69	6	—	6	2
B	1.70–2.09	16	—	18	4
C	2.50–2.79	8	4	8	4
D	2.80–3.09	6	6	6	6
E	3.50–3.79	4	8	4	8
F	3.80–4.29	—	16	4	18
G	4.40–4.79	—	6	2	6

[a] Baseline ratings taken from Geiselman et al. (1984).

approximately equal weighting of range and frequency tendencies (Birnbaum, 1974; Parducci, 1963; Parducci & Perrett, 1971; Riskey, Parducci, & Beauchamp, 1979). However, under certain conditions the best-fit frequency weighting, $1 - w$, has varied inversely with the number of rating categories (Parducci, 1982; Parducci & Wedell, 1986).

Experiment 1 was designed to test the fit of range-frequency theory to judgments of physical attractiveness. Positively and negatively skewed distributions were created by varying either frequencies or both range and frequencies of different faces presented by the method of single stimuli. To determine whether sensitivity to the skewing of contextual sets would diminish with more categories, the number of permissible rating categories was also varied.

Method

Design

The experiment used a four-way factorial design, with Skewing (positive and negative), Range (full and restricted), Number of Categories (5, 9, and 101), and Stimulus (three stimulus groupings common to all sets) as independent factors. Stimulus was the only within-subject factor. The dependent variable was the rating assigned to each of the test faces common to all four distributions.

Stimuli

The stimuli were drawn from a set of 118 photographic slides of female faces taken from a high school year book. The faces were divided into seven groups on the basis of earlier ratings using a five-point scale (Geiselman et al., 1984). Table 1 shows how each group was defined and also the relative frequencies of the groups for different contextual conditions. To analyze the effects of context, certain stimuli were common to both positive and negative sets. There were six of these common stimuli for the restricted-range conditions, two each from groups C, D, and E. There were 10 common stimuli for full-range conditions, two each from groups B–F.

Presentation sequences were determined by a block-randomization procedure: two blocks of 20 for restricted-range sets, two blocks of 24 for full-range sets. The test stimuli appeared in the same serial positions for corresponding positive and negative sets. For each condition, the slides were presented in forward before reverse order for approximately half the subjects, but reverse before forward order for the other half.

Procedure

Subjects were instructed to rate the attractiveness of each face using either a 5-, 10-, or 101-point numerical scale. For each scale, the lowest number was labeled *Very Unattractive* and the highest *Very Attractive*. The middle category for the 5- and 101-point scales was labeled *Neutral*. The actual numbers on the scale ranged from 1 to 5, 1 to 10, and −50 to 50. Subjects were instructed to record numerical ratings. They were told that although faces would be repeated, there was no need to remember their previous ratings but to record just their immediate impressions. The faces were projected every 5 s, with a 0.5-s switching interval onto a screen 1–3 m from the subjects.

Subjects

A total of 281 students participated to fulfill a requirement for introductory psychology at the University of California, Los Angeles. From 20 to 28 were assigned randomly to each skewing × range × category condition. Testing was in groups of 8–12.

Results

The top six panels of Fig. 1 plot the mean rating of each group of stimuli for the four contextual conditions. The differences between the two rating functions in each panel reveals the expected contrast effect: Ratings for the same test faces are higher when contextual stimuli are predominantly unattractive. The effects of context upon ratings of the common faces (viz., Groups C, D, and E) are approximately twice as great for the restricted range (top row) as for the full range (middle row). Comparing across columns, the number of categories has little effect on the magnitude of the context effect; however, there is a tendency for stimuli to be rated higher as the number of categories increases.

A four-way, analysis of variance (ANOVA) was performed upon the mean ratings of the test stimuli from Groups C, D, and E, linearly transformed to a common scale using the following equation:

$$J_{ic} = (C_{ic} - a)/b \qquad (2)$$

where J_{ic} is the transformed judgment (on a scale from 0 to 1) of Stimulus i in Context c, a is the number corresponding to the lowest category and b is the permissible range of the scale (e.g., four for a five-point scale).

The contrast effect of Skewing was highly significant, $F(1, 269) = 88.86$, $p < .001$. The significant Skewing × Range interaction, $F(1, 269) = 5.69$, $p < .05$, reflects the greater contrast effect for the restricted-range conditions. The main effect of Category was also significant, $F(2, 269) = 31.85$, $p < .001$, reflecting an overall increase in mean ratings with more categories. Although the interaction between Skewing and Number of Categories was not significant ($F < 1$), the Category × Range

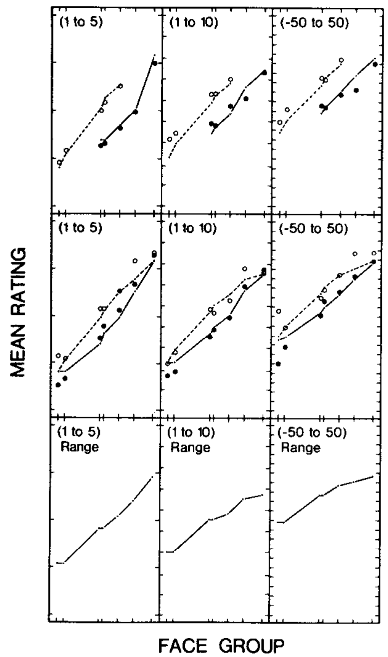

FIG. 1. Contrast effect for ratings of attractiveness. Open and closed circles represent mean ratings for positively and negatively skewed sets, respectively; lines represent theoretical fit of the range-frequency model with $1 - w = 0.24$ for all conditions and inferred range values as shown in the bottom panels. Spacing of the face groups (A–G) is proportional to their differences in mean ratings across all conditions.

interaction was, $F(2, 269) = 4.25$, $p < .05$. Each successive increase in number of categories resulted in a significant increase in overall mean rating for restricted-range conditions; although these increases were smaller for full-range conditions, all but the difference separating 5- and 10-point means was significant.

The only other effect to reach statistical significance was the main effect of Stimulus, $F(2, 538) = 83.78$, $p < .001$, indicating that subjects distinguished among the three test groups. Only the linear component of this effect was statistically significant ($p < .001$).

Range-Frequency Fits

To test how well range-frequency theory fits the data, frequency values were first calculated using the counting algorithm employed in previous applications of the theory (Parducci & Perrett, 1971). The total number of experimental presentations was first divided by the number of categories to obtain the number of presentations that would have been assigned to

each category if subjects were using categories equally often (while keeping a perfectly ordinal scale). For example, when the 48 presentations of the full-range condition are rated using five categories, there would be 9.6 presentations assigned to each category on the average; thus for the negatively skewed distribution, faces in Groups A and B (see Table 1) would receive a rating of 1 (on a 5-point scale), since their cumulative frequencies are less than 9.6. For faces in Group C, 3.6 of the presentations would be assigned a rating of 1 and 0.4 presentations a rating of 2 for a combined mean rating of 1.1. Continuing with this procedure, frequency values for faces in Groups D, E, F, and G would be 2.0, 2.6, 3.9, and 5.0, respectively. An advantage of the counting procedure is that it allows for some of the stepwise character of rating functions found when the number of categories is relatively small; the algorithm is not meant to be descriptive of the psychological processes underlying the frequency principle.

Transposing Eq. (1) for each distribution and subtracting yields

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

where the plus and minus subscripts indicate the positively and negatively skewed contexts, respectively. When the extreme stimuli are the same, range values for positive and negative sets may be assumed to be equal and so drop out; thus for the full-range conditions

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

A single weighting value was determined for each number of rating categories (5, 10, 101) by averaging the $1 - w$ values calculated for the five common test groups (B–F) of the full-range condition. If the subjective range is defined by the endpoints of the experimental sets, then application of Eq. (4) to the restricted-range sets should overestimate $1 - w$, since differences in range values will contribute to the contrast effect. To test whether the subjective range varied with the different experimental ranges, Eq. (4) was applied to the three common test groups (B, C, and D) of the restricted-range sets to estimate $1 - w$ for corresponding 5-, 10-, and 101-point scales. Because these estimates of $1 - w$ were virtually the same as those for the full-range sets, range values were assumed *not* to vary with experimental manipulations of range. The six values of $1 - w$ were averaged together to arrive at a single value, 0.24, for all conditions.

This value for $1 - w$ was then substituted into Eq. (1) along with frequency values and empirical ratings to infer range values. Although range values were assumed not to vary with context or range, they were assumed to vary with number of categories (as implied by the highly significant main effect of categories). Thus, a separate range function was inferred for each number of categories by averaging together the range values calculated from Eq. (1) for the different contextual conditions (with the restriction that the function remain monotonic). These are shown in the bottom row of Fig. 1.

The theoretical functions in the top two rows of Fig. 1 were generated by substituting the inferred range values, 1 − *w* (0.24), and predetermined frequency values into Eq. (1). Although the fit is not as good as similar fits to psychophysical data (cf. Parducci & Wedell, 1986), the range-frequency model captures the general pattern of the data surprisingly well (correlation between theoretical and data points is .968). Some of the larger errors in predicted values may be attributed to the large individual differences in judgment. There were many instances where a face was rated highly attractive by one subject and below average in attractiveness by a different subject in the same condition. It should be emphasized that the 12 rating scales are fit with no parameters varying with skewing or range and only the parameters defining the range function varying with number of categories.

Discussion

The range-frequency model accounts for the contrast effect between positive and negative sets and also for contrast being greater when the stimulus range is restricted. In experiments with simple psychophysical stimuli, the subjective range is typically defined by the particular end stimuli of the presented series. However, the range-frequency analysis of the data indicates that the greater contrast for restricted-range sets was due to greater corresponding differences in frequency values between positive and negative sets rather than a systematic shift in range values. In rating the physical attractiveness of the present set of faces, subjects appear to invoke reference stimuli more extreme than any of those actually presented.

One of the most striking aspects of the data is the relatively higher ratings assigned faces when more categories were made available. For example, when the faces in Group F were presented in the restricted-range, negatively skewed context, they were rated slightly below average on a 5-point scale but above average on 10- and 101-point scales. The difference in mean ratings between 5- and 10-point scales indicates that this effect is not restricted to the 101-point scale or the use of negative numbers (−50–50).

The range-frequency analysis attributes these differences to shifts in the underlying range functions. The decreasing slope and higher *Y* intercept of the range function with more categories implies an increase in the tendency to invoke faces of lower attractiveness, such as grotesque horror masks, as reference stimuli. This tendency may be motivated by the communicative intention implied by the number of rating categories. When subjects are limited to just a few categories, they may feel compelled to use all of the available categories to communicate the differences they

perceive within the stimulus set. However, presenting subjects with a large number of rating categories may encourage them to retrieve stimuli more extreme than those presented in the experimental set and include these in the effective context for judgment. The set of faces used in this experiment (yearbook pictures of high school seniors) seems to fall within the moderate to attractive range of possible experience, and so the subjective range was extended downward to include less attractive faces. If the set of experimental faces were predominantly unattractive, one would expect an upward extension of the subjective range and hence lower ratings with more categories. A similar tendency to extend the subjective range beyond the stimulus set with increasing number of categories was found for ratings of sizes of squares presented in narrow contextual ranges spanning only a few discriminable values (Parducci, 1982).

Although increasing the number of categories results in a systematic shift in the range values, it does not affect the relative weighting of range and frequency principles. The category effect (viz., the reduction in skewing effects with more categories) occurs when skewing is achieved by presenting stimuli with *unequal frequency* (Parducci, 1982); it does not occur when skewing is achieved using unequally spaced stimulus values presented with equal frequency (Parducci & Wedell, 1986; Mellers & Birnbaum, 1982). In the present experiment, groups of stimuli were presented with unequal frequency in an attempt to produce a category effect. However, each group consisted of several *different* faces. Memory for individual faces is extremely accurate (for a review, see Davies, Shepherd, & Ellis, 1981), and thus the effective context may have been characterized by unequal spacing rather than unequal frequencies. If so, the category effect should not have been expected.

The frequency weighting of 0.24 is only half that typically obtained in psychophysical experiments. This may be due to greater weighting of range values for judgments of familiar, easily recognized stimuli. Another possible interpretation is that range and frequency values are weighted equally, with another factor, the responses assigned to similar stimuli in the past, also weighted into the judgment. Incorporating values from a previously established response scale helps stabilize the scale and may enhance the transmission of information (Wedell & Parducci, 1985). The importance of stimulus–response associations in judgment is exemplified in the "change-of-standard" effect (Higgins & Lurie, 1983). Wedell (1984/1985) utilized a response-averaging model to explain the reduction in contrast effects that occurs following a shift in the contextual set for judgment. However, the present experiment cannot differentiate between response-averaging and change-in-weighting interpretations.

EXPERIMENT 2: COMBINING CONTRAST AND ASSIMILATION EFFECTS USING PAIRED STIMULI

Experiment 1 explained contrast as a range-frequency effect of contextual distribution. Assimilation, on the other hand, may be modeled as an averaging effect. Integration theory (Anderson, 1965, 1981) describes the judgment of a stimulus as a weighted average of its value in isolation and the value of its associated context. This integration of stimuli may be represented algebraically as follows:

$$S_{ic} = zS_i + (1 - z)S_c \tag{5}$$

where S_{ic} is the subjective value of Stimulus i in Context c, S_i is its value when stimuli are presented successively, S_c is the corresponding value of simultaneously presented contextual stimuli, and z is the weighting parameter. Thus, according to the integration interpretation, the assimilation effect results from a failure to completely separate the individual stimulus from other stimuli in its immediate contextual set.

In support of Eq. (5), Anderson, Lindler, and Lopes (1973) found that hypothetical persons characterized by brief verbal descriptions were rated more attractive when represented as members of an attractive rather than an unattractive group. This finding has been extended to ratings of physical attractiveness by several other researchers (Geiselman et al., 1984; Melemed & Moss, 1975; Strane & Watts, 1977). An important finding of the Geiselman et al. study was that assimilation occurred only when faces were presented simultaneously; a trend toward contrast rather than assimilation was found when the faces were presented successively.

If contrast and assimilation are related to different arrangements of the context, it should be possible to demonstrate both effects within the same experiment and thus study how range-frequency and integration theories can best be combined to explain joint effects. Experiment 2 was designed to do this. The distribution of contextual stimuli consisted mostly of either high-, moderate-, or low-attractive faces, creating negatively skewed, bell-shaped, or positively skewed distributions, respectively. In accordance with the range-frequency interpretation of Experiment 1, the distribution of successive stimuli was expected to produce a contrast effect between sets: moderate test faces to be rated highest for the positive set, lowest for the negative set.

Experiment 2 differed from Experiment 1 in that faces were presented in pairs rather than singly. Moderate test faces were presented simultaneously with either attractive or unattractive faces. In accordance with the findings of Geiselman et al. (1984), an assimilation effect was expected within pairs: A moderate face should be rated more favorably when presented alongside a more attractive face. Thus, this experiment is designed to produce both contrast and assimilation effects for ratings of the same stimuli at the same time. The use of three distributions facilitates testing between different models of how judgmental processes producing assimilation are combined with those producing contrast.

Method

Design

The experiment was a three-way, factorial design, with between-pair context (positively skewed, negatively skewed, and bell-shaped distributions), within-pair context (low and high), and number of categories (5 and 101) as independent variables. Within-pair context was the only within-subject factor. The dependent variable was the mean rating of the moderate test stimuli common to all conditions.

Stimuli

Stimuli were drawn from the same set of slides sampled for Experiment 1. Each experimental set consisted of 12 pairs. The two slides constituting a pair were projected simultaneously on the screen (by two projectors). The first six pairs of faces were either of high (H) attractiveness, selected from Groups F and G, moderate (M) attractiveness, selected from Group D, or low (L) attractiveness, selected from Groups A and B. These six contextual pairs were followed by six test pairs. Each test pair consisted of a context face, projected on the left side of the screen, and a test face, projected on the right. All six test stimuli were M faces. Context stimuli for three of the test pairs were H faces, the other three were L faces.

The order of the six test faces was the same for all conditions. To counterbalance effects specific to particular faces, each test face appeared equally often with an L and an H face. To counterbalance order effects, the two types of context faces were alternated on successive trials, beginning either with an H face or an L face.

Procedure

To facilitate assimilation, subjects were told that each pair of faces represented "friends from the same high school." They were instructed to judge the attractiveness of each face separately, rating the face on the left first, then the face on the right. Ratings were made using either a 5- or 101-point scale. In all other ways the instructions and the labeling of the rating scales were the same as those described in Experiment 1. Each pair of slides was projected for 14 s, with a 0.5-s interstimulus interval. Once again, subjects were tested in groups of 8–12.

Subjects

Subjects were 251 students sampled from the same source as in Experiment 1. Approximately 40 subjects were assigned randomly to each category × skewing condition.

Results

For each panel of Fig. 2, the vertical difference between the solid-line functions indicate within-pair assimilation, the one exception to this trend occurring for the test faces at the five-category, low between-pair context. The downward slopes associated with differences in between-pair context indicate a contrast effect which is greater for the 5-point scale. Faces are rated much more positively on the 101-point scale than on the 5-point scale.

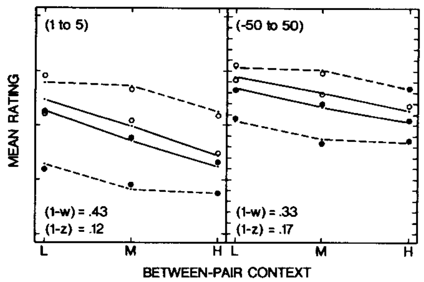

FIG. 2. Assimilation and contrast for ratings of attractiveness Theoretical fit of the combined range-frequency/integration model to test faces (solid lines) and contextual faces (dashed lines). Assimilation effect represented by higher ratings when presented with high within-pair context (open circles) vs low within-pair context (closed circles). Contrast effect represented by decrease in ratings across between-pair contexts, low (L), medium (M), and high (H).

A three-way ANOVA was conducted on the ratings of the test stimuli (after linear transformation to a common scale using Equation (2)). The main effect of Within-Pair Context was statistically significant, $F(1, 239) = 26.63$, $p < .001$, confirming the assimilation effect. The main effect of Between-Pair Context was also significant, $F(2, 239) = 36.64$, $p < .001$, confirming the contrast effect. Their interaction was not significant, $p > .05$.

The main effect of Number of Categories was again highly significant, $F(1, 239) = 84.03$, $p < .001$, reflecting the higher ratings given in the 101-point condition. The interaction between the effects of Between-Pair Context and Number of Categories approached significance, $F(2, 239) = 2.77$, $.10 > p > .05$. A more sensitive test of this effect is given by the linear component of the interaction, which did achieve statistical significance, $F(1, 239) = 3.94$, $p < .05$, consistent with the category effect reported by Parducci (1982). None of the other tests was statistically significant.

Theoretical Fits

Range-frequency theory (Eq. (1)) was used to fit the effects of between-pair context, and integration theory (Eq. (5)) was used to fit the effects of within-pair context. Tests were made of three different models of how these two theories combine. These models differ in the sequencing of the psychological processes corresponding to the two theories. One model assumes that within-pair and between-pair contextual processing take place in *parallel,* the overt rating representing a weighted average of the results of the two processes. A second model assumes that within-pair contextual processing occurs first and that the resulting values, the S_{ic} of Eq. (5), represent the range values, the R_{ic} of Eq. (1), in the range-frequency processing of the between-pair context. The third model reverses the order of the two processes: Judgments (C_{ic} of Eq. (1)) resulting from the range-frequency analysis of each member of the pair represent the values (S_i and S_c) used in the within-pair averaging process.

Parallel processing. This model requires that the results of range-frequency and integration processes be made commensurable before they are averaged together. To accomplish this, the stimulus values of Eq. (5) (S_i and S_c) must be represented by their corresponding range values (used also in Eq. (1)). The magnitudes of contrast and assimilation effects then are given by $1 - w$ of Eq. (1) and $1 - z$ of Eq. (5), respectively (when w and z are both 1 neither contrast nor assimilation is observed). Because the range value of the stimulus being judged is used in both between-pair and within-pair contextual processing, it receives greater weighting. If the results of the two processes were actually weighted equally, then $1 - w$ would have to be quite large (0.86 and 0.66 for 5- and 101-point scales) in order to fully capture between-pair differences. However, doing so yields inferred range values (from Eq. (1)) that differ greatly for the different distributions (i.e., by more than the range of categories) and hence provide a poor fit to the data. To reduce differences in range values inferred for the different between-pair contexts requires a large reduction in the weighting for frequency values. But this precludes satisfactory fits of the between-pair contextual effects.

Within-pair processing first. This model assumes that the results of the integration process (Eq. (5)) are used to determine range values for subsequent range-frequency analysis (Eq. (1)). Thus, the same face would have a higher range value when paired with an attractive than with an unattractive face. But increasing the range value of the face would produce a corresponding increase in its rank or frequency value in the distribution of contextual stimuli (i.e., the number of faces with lower range values would increase). This difference in the frequency values for the same test face when paired with a high vs low contextual face means that the range-frequency process would itself enhance the assimilation effect, the enhancement increasing with greater differences in frequency values.

The three distributions used in the present experiment provide a crucial test of this model. Specifically, the model predicts that the assimilation effect for the bell-shaped distribution will be more than five times greater than that for either of the skewed distributions (due to the greater shift in frequency values for M test stimuli with the bell set). This prediction is clearly falsified by the data. In order to get a reasonable fit to the assimilation effect, the model requires a very low value of $1 - w$, but this results in a failure to fit differences between distributions.

Range-frequency processing first. The model that fits the data best puts the range-frequency evaluation first, using the resulting values of judgments as stimulus values for the within-pair integration process.

Figure 2 shows the fit of this model to the data. Only 4 of the 24 theoretical points err by more than one standard error of the corresponding mean rating. Figure 2 also shows the best-fitting values for $1 - w$ and $1 - z$, the respective weightings of between-pair and within-pair contexts. These values plus the range value for each of the three stimulus groups (L, M, and H) were inferred for each number of categories using predetermined frequency values. The inferred range values for the L, M and H faces were 2.17, 2.78, and 2.97 for five categories and 6.89, 18.97, and 19.84 for the -50 to 50 scale. These range values are similar to the range values inferred for Experiment 1 (see Fig. 1). The obtained frequency weighting, $1 - w$, is considerably greater for five categories, but the contextual weighting factor, $1 - z$, is only slightly greater for the 101-point scale.

Discussion

The assumption that a range-frequency compromise precedes the integration of stimuli yields a good account of both the contrast and assimilation found in Experiment 2. Stimuli are first located within the contextual distribution by range-frequency processing and then an integration process averages these locations for the two members of the presented pair. Neither of the other models (parallel or within-pair processing first) provides an adequate fit to the data.

Although range-frequency analysis appears to take precedence over the integration process in judgments of physical attractiveness, this may not be the case when the paired stimuli are more integrally related (e.g., judging a rectangle's height and width or a person's intelligence and gullibility). When components are more completely integrated, one might expect a reversal in the sequencing of the two processes.

As in Experiment 1, increasing the number of categories resulted in higher ratings of attractiveness, as though the subjective range were extended to include faces more unattractive than any of those presented in the experimental series. Unlike Experiment 1, but in line with the category effect found with psychophysical stimuli (Parducci, 1982), the effects of contrast were greater for 5- than for 101-point scales. There are several procedural differences between Experiments 1 and 2 that might be responsible for this difference (e.g., number of faces presented, number of different face groups, duration of presentation, random vs ordered sequence, single vs dual presentation). In particular, the initial massed presentation of faces of similar value (the first six pairs were all H, L, or M faces) may have encouraged subjects to group these together, creating the unequal frequency distribution necessary for the category effect to occur (Parducci & Wedell, 1986). The present results demonstrate that the category effect is not restricted to psychophysical stimuli. Indeed, we have also observed it for ratings of the happiness of life events as

described in short sentences (Wedell & Parducci, 1986) and ratings of the mental disturbance portrayed in clinical case histories (Wedell, Parducci, & Lane, 1986).

DISTINGUISHING AMONG ALTERNATIVE INTERPRETATIONS

Because contrast and assimilation represent the only possible directions of contextual shift, it seems likely that these effects may result from a number of different cognitive processes. Although the theoretical fits to the data of Experiments 1 and 2 support the application of both range-frequency and integration theories to explain how the different kinds of context affect ratings of physical attractiveness, there are a number of plausible alternative interpretations. In this section, some of the more common theoretical interpretations of contrast and assimilation are discussed in relation to the present data.

Contrast Effects

Contrast is generally described as the result of a process of comparison—to a single standard representing a neutral point, as in Helson's (1964) adaptation-level theory or Thibaut and Kelley's (1959) comparison-level theory, to extreme end stimuli anchoring the range, as in Volkmán's (1951) range theory, or to both the range and ranks of contextual stimuli, as in Parducci's (1963) range-frequency theory. All of these theories describe contrast as an effect of the contextual distribution and all can provide adequate accounts of the contrast effects observed here.

The single-standard theories generally locate the neutral point of the scale at the mean of the contextual distribution. Judgments then reflect deviations from this neutral standard. Thus, when attractive faces predominate in the series, a face of average attractiveness will fall below the mean and receive a low rating. Restricting the lower range for the attractive context raises the mean and results in even lower ratings.

According to range theory, judgments reflect the location of the stimulus within a range of values determined by the contextual stimuli. When the range is restricted to attractive faces, a face of average attractiveness will fall in the lower half of the range and receive a low rating. However, when the stimulus range is equated for attractive and unattractive contexts (as in the full-range conditions of Experiment 1), range theory requires that the subjective range extend beyond the more frequently presented range anchor in order to account for the contrast effect.

Although the present experimental results do not distinguish among these different theories, a number of crucial tests have supported the range-frequency account (Birnbaum, 1974; Parducci, 1963; Parducci & Perrett, 1971). The unique feature of range-frequency theory is that judgment are sensitive to the relative densities of the stimuli along the range: The widths of rating categories decrease where stimuli are densely packed, possibly reflecting an intention to communicate perceived differences among the stimuli. Thus, range-frequency theory allows for the "contrast effect" to occur in different directions within the same stimulus distribution. For example, when stimuli are densely packed at either end of the range (a *U*-shaped distribution), ratings of moderately positive stimuli are displaced downward, while at the same time, ratings of moderately negative stimuli are displaced upward. Because adaptation-level and range theories presume that ratings are linearly related to underlying scale values, they cannot account for the nonlinear shifts that result from manipulating the relative densities of stimuli along the range.

The conception of contrast as a distribution effect may appear at odds with its occurrence when only a few stimuli are presented. For example, Kenrick and Gutierres (1980) found that ratings of the physical attractiveness of a target face were shifted significantly in the direction of contrast when preceded by presentation of a single contextual photograph. These authors argued that theories of scale usage, such as range-frequency theory, could not account for the observed contrast effect, since only a single face was actually rated. Although range and frequency principles were originally described in terms of how subjects use their response categories (Parducci, 1963, 1965), they may also be described without reference to category usage (Birnbaum, 1974; Parducci, 1983). For example, the frequency principle may be described either as a tendency to use response categories with equal frequency or as a tendency to judge the stimulus on the basis of its rank within the contextual distribution. Both formulations entail that judgments reflect the ranks of the stimuli, but they imply different processes underlying this effect—one bound to category usage and the other not. That contrast effects occur across a variety response modes, such as magnitude estimations (Mellers & Birnbaum, 1982), matching tasks (Manis, 1967), and open-ended verbal descriptions (Simpson & Ostrom, 1976) supports the conception of a response-independent effect of context. On the other hand, the recent finding that the effects of skewing contextual frequencies decreases with more categories (Parducci & Wedell, 1986) suggests that response categories may play some direct role in the processing of contextual stimuli.

In either case, range-frequency theory assumes that judgments are always made relative to a distribution of contextual stimuli. Presentation of a stimulus to be judged along a particular dimension is assumed to trigger automatic retrieval of a distribution of comparison stimuli as described recently by norm theory (Kahneman & Miller, 1986). The judgment of the stimulus then represents its location within this distribution according to range-frequency principles. In this way, presentation of even a single contextual stimulus may alter the distribution by either extending the range or increasing the density within a region of the range. In Experiment 1, it appears that contrast was due to increased

densities rather than extensions of the range. However, increasing the number of categories in the rating scale appears to extend the range of stimuli included in the context for judgment.

Assimilation Effects

Several theoretical explanations of assimilation are described in the current literature. Recent work by Manis and his associates (Manis & Paskewitz, 1984; Manis, Paskewitz, & Cotler, 1986) has focused upon an expectation-driven basis for assimilation: The subject tends to judge the stimulus in accordance with expectations generated through previous experience with contextual stimuli. The tendency to match responses to expectations has been used to explain the basic phenomenon of probability matching (Estes, 1972) as well as the "assimilative" shift in the judgment criterion when stimulus frequencies are varied in signal-detection tasks (Green & Swets, 1966) or in pairwise discrimination tasks with feedback (Parducci & Sandusky, 1970). Within this framework, the assimilation effect of Experiment 2 could be explained in terms of an expectation that faces within a pair will be of similar attractiveness. Such an expectation could have resulted from the fact that faces in the initial six pairs were matched in value (all of high, moderate, or low attractiveness) or it could simply be the result of the designation of individuals within pairs as "friends." However, Geiselman et al. (1984), using these same stimuli and procedures, found strong assimilation effects with pairs that were predominantly mixed (rather than matched) in value and regardless of whether pairs were described as friends or not. Thus, expectation-driven assimilation seems less applicable to the present data set.

Assimilation effects may also be driven by a process of cognitive inference. For example, Sigall and Landy (1973) proposed an attribution-based explanation of why an individual is evaluated more favorably when associated with an attractive rather than an unattractive partner. The reasoning behind the "radiating beauty" effect is that in order for a person of average attractiveness to become associated with an attractive partner, that person must presumably have many other favorable attributes. However, this interpretation predicts that judgments of physical attractiveness will not shift; rather, it locates the assimilative effect along other dimensions of judgment in order to compensate, in effect, for a lack of a match in physical attractiveness. Consistent with this prediction, Sigall and Landy found no assimilation effect for judgments of physical attractiveness in their data.

A third class of interpretations proposes that contextual stimuli result in a "change in meaning" of the target stimulus (Hamilton & Zanna, 1974; Wyer, 1974). This type of interpretation has gained prominence within the recent processing-oriented approach to social cognition (Higgins, Rholes, & Jones, 1977; Srull & Wyer, 1980). Presentation of contextual stimuli primes a set of associated semantic nodes. Because primed nodes are more readily activated than unprimed nodes, the retrieved meaning of the target stimulus will tend to be consistent with the meaning of contextual stimuli. Herr, Sherman, and Fazio (1983) have demonstrated that ambiguity of the target stimulus and the relative proximity of its value to contextual stimuli are both necessary conditions for assimilation to operate within this framework. Only when the target is ambiguous (i.e., has several alternative meanings) and relatively close in value to contextual stimuli can it share connections through associated nodes and hence shift its meaning assimilatively.

For judgments of physical attractiveness, nodes may be conceived as featural (hairstyle, shape of nose, etc.) rather than semantic in nature. Faces may then be considered "ambiguous" in that they share many features. Accordingly, the shared features of simultaneously present faces will receive greater activation and thus shift the perceived values of the faces closer together. This interpretation further predicts that when the difference in attractiveness levels of target and contextual stimuli exceeds some critical value, contrast rather than assimilation should be observed (Herr et al., 1983). However, one difficulty in applying this interpretation to the present data is that contrast was observed when faces were presented successively (Experiment 1). With an interstimulus interval of only 0.5 s, priming from the previous trial would be expected to persist through the subsequent trial. Since either attractive or unattractive faces predominated in the series, an overall assimilation effect (rather than contrast) would have been predicted.

A fourth class of models locates the assimilation effect at the integration rather than encoding or retrieval stages of processing. Anderson's (1981) averaging formulation (Eq. (5)) exemplifies this approach. The model assumes that scale values of the stimuli do not change; instead, the values of contextual stimuli are averaged with the value of the target stimulus to arrive at the psychological impression. In its strictest form, the weighting value is independent of stimulus values and so the model predicts parallel effects across the range of values.

One way to conceive of the process underlying the averaging model is in terms of an anchoring and adjustment strategy (Lopes, 1981). For example, the subject may anchor the judgment scale on the overall impression of the pair of faces and then adjust the judgment toward the value of the individual face being rated. Because adjustment is typically incomplete (Tversky & Kahneman, 1974), assimilation is observed. The weighting of the individual stimulus in Eq. (5) may be interpreted as reflecting the degree of adjustment toward the stimulus value (with $z = 1$ representing complete adjustment). This version of the anchoring and adjustment formulation entails an assimilation effect on judgments of both target and contextual stimuli. Alternatively, if the more extreme contextual stimulus were to serve as the anchor, with adjustment directed toward the moderate target stimulus, assimilation would occur only for the target.

Because these models are descriptive in nature, conditions under which assimilation occurs must be determined empirically. The work of Geiselman et al. (1984) indicates that simultaneous presentation is the main determinant of the assimilation effect for judgments of physical attractiveness. However, the generality of the simultaneous assimilation effect is unclear. For example, the number of simultaneously present stimuli appears to be a crucial consideration. When several stimuli are presented on a page, contrast rather than assimilation is observed (Mellers & Birnbaum, 1982; Parducci & Wedell, 1986). One possibility is that there is assimilation to immediately adjacent stimuli but contrast to the overall distribution. But this type of local assimilation appears to depend on the nature of the stimuli. For example, the well-known Ebbinghaus illusion clearly demonstrates contrast to adjacent stimuli (Massaro & Anderson, 1971). More recently, Russell (1985) has found a contrast effect for judgments of emotion portrayed in faces when presented in pairs. Thus, conditions under which simultaneous presentation results in assimilation rather than contrast have yet to be clearly outlined.

The foregoing discussion points to a number of plausible processes underlying assimilation. We argue that the simultaneous assimilation observed for judgments of physical attractiveness may be best conceived as an averaging process occurring in the integration stage. More generally, however, assimilation may result from any of several cognitive processes. One goal for future research is to more carefully determine the domain of applicability for each of the various theories.

REFERENCES

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

Anderson, N. H. (1981). *Foundations of integration theory.* New York: Academic Press.

Anderson, N. H., Lindler, R., & Lopes, L. L. (1973). Integration theory applied to judgments of group attractiveness. *Journal of Personality and Social Psychology,* **26,** 400–408.

Beebe-Center, J. G. (1932). *Pleasantness and unpleasantness.* New York: Van Nostrand.

Birnbaum, M. H. (1974). Using contextual effects to derive psychophysical scales. *Perception & Psychophysics,* **15,** 89–96.

Davies, G. M., Shepherd, J. W., & Ellis, H. D. (1981). *Perceiving and remembering faces.* New York: Academic Press.

Estes, W. K. (1972). Research and theory on the learning of probabilities. *Journal of the American Statistical Association,* **67,** 81–102.

Geiselman, R. E., Haight, N. A., & Kimata, L. G. (1984). Context effects on the perceived physical attractiveness of faces. *Journal of Experimental Social Psychology,* **20,** 409–424.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics.* New York: Wiley.

Hamilton, D. L., & Zanna, M. P. (1974). Context effects in impression formation: changes in connotative meaning. *Journal of Personality and Social Psychology*, **29**, 649–654.

Helson, H. (1964). *Adaptation-level theory*, New York: Harper & Row.

Herr, P. M., Sherman, S. J., & Fazio, R. H. (1983). On the consequences of priming: Assimilation and contrast effects. *Journal of Experimental Social Psychology*, **19**, 323–340.

Higgins, E. T., & Lurie, L. (1983). Context, categorization, and memory: The "change-of-standard" effect. *Cognitive Psychology*, **15**, 525–547.

Higgins, E. T., Rholes, W. S., & Jones, C. R. (1977). Category accessibility and impression formation. *Journal of Experimental Social Psychology*, **13**, 141–154.

Kahneman, D., & Miller, D. T. (1986). Norm theory: Comparing reality to its alternatives. *Psychological Review*, **93**, 136–153.

Kenrick, D. T., & Gutierres, S. E. (1980). Contrast effects and judgments of physical attractiveness: When beauty becomes a social problem. *Journal of Personality and Social Psychology*, **38**, 131–140.

Kernis, M. H., & Wheeler, L. (1981). Beautiful friends and ugly strangers: Radiation and contrast effects in perceptions of same-sex pairs. *Personality and Social Psychology Bulletin*, **7**, 617–620.

Lopes, L. L. (1981). *Averaging rules and adjustment processes: The role of averaging in inference* (Rep. 13). Madison: University of Wisconsin, Wisconsin Human Information Processing Program.

Manis, M. (1967). Context effects in communication. *Journal of Experimental Psychology*, **5**, 325–234.

Manis, M., & Paskewitz, J. (1984). Judging psychopathology: Expectation and contrast. *Journal of Experimental Social Psychology*, **20**, 217–230.

Manis, M., Paskewitz, J., & Cotler, S. (1986). Stereotypes and social judgment. *Journal of Personality and Social Psychology*, **50**, 461–473.

Massaro, D. W., & Anderson, N. H. (1971). Judgmental model of the Ebbinghaus illusion. *Journal of Experimental Psychology*, **89**, 147–151.

Melemed, L., & Moss, M. K. (1975). The effect of context on ratings of attractiveness of photographs. *Journal of Psychology*, **90**, 129–136.

Mellers, B. A., & Birnbaum, M. H. (1982). Loci of contextual effects in judgment. *Journal of Experimental Psychology: Human Perception and Performance*, **8**, 582–601.

Parducci, A. (1963). Range-frequency compromise in judgment. *Psychological Monographs*, **77**, (92, 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*. Hillsdale, NJ: Erlbaum.

Parducci, A. (1983). Category ratings and the relational character of judgment. In H. G. Geissler, H. F. J. M. Buffort, E. L. J. Leeuwenberg, & V. Sarris (Eds.), *Modern issues in perception* (pp. 262–282). Berlin: VEB Deutsher Verlag der Wissenschaften.

Parducci, A., Perrett, L. F. (1971). Category rating scales: Effects of relative spacing and frequency. *Journal of Experimental Psychology Monograph*, **89**, 427–452.

Parducci, A., & Sandusky, A. J. (1970). Limits on the applicability of signal detection theories. *Perception & Psychophysics*, **7**, 63–64.

Parducci, A., & Wedell, D. H. (1986). The category effect with rating scales: Number of categories, number of stimuli, and method of presentation. *Journal of Experimental Psychology: Human Perception and Performance*, **12**, 496–516.

Riskey, D. R., Parducci, A., & Beauchamp, G. K. (1979). Effects of context in judgments of sweetness and pleasantness. *Perception & Psychophysics*, **26**, 171–176.

Russell, J. A. (1985). *Relativity in the perception of emotion via facial expression*. Unpublished manuscript.

Sigall, H., & Landy, D. (1973). Radiating beauty: Effects of having a physically attractive partner on person perception. *Journal of Personality and Social Psychology*, **28**, 218–224.

Simpson, D. D., & Ostrom, T. (1976). Contrast effects in impression formation. *Journal of Personality and Social Psychology*, **34**, 625–629.

Srull, T. K., & Wyer, R. S., Jr. (1980). Category accessibility and social perception: Some implications for the study of person memory and interpersonal judgments. *Journal of Personality and Social Psychology*, **38**, 841–856.

Strane, K., & Watts, C. (1977). Female judged by attractiveness of partner. *Perceptual and Motor Skills*, **45**, 225–226.

Thibaut, J. W., & Kelley, H. H. (1959). *The social psychology of groups*. New York: Wiley.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, **185**, 1124–1131.

Volkman, J. (1951). Scales of judgment and their implications for social psychology. In J. Rohrer & M. Sherif (Eds.), *Social psychology at the crossroads* (pp. 273–294). New York: Harper & Row.

Wedell, D. H. (1985). A process model for psychophysical judgment (Doctoral dissertation, University of California, Los Angeles, 1984). *Dissertation Abstracts International*, **45**, 3102B.

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). New York: Elsevier.

Wedell, D. H., & Parducci, A. (1986). *The category effect in social judgment: Experimental ratings of happiness*. Unpublished manuscript.

Wedell, D. H., Parducci, A., & Lane, M. (1986). *Reducing contextual effects for ratings of clinical case histories*. Unpublished manuscript.

Wyer, R. S., Jr. (1974). Changes in meaning and halo effects in personality impression formation. *Journal of Personality and Social Psychology*, **29**, 829–835.