JOURNAL OF EXPERIMENTAL PSYCHOLOGY
MONOGRAPH

Vol 102, No 3, 543–561 March 1974

# THE NONADDITIVITY OF PERSONALITY IMPRESSIONS [1]

MICHAEL H BIRNBAUM [2]

*University of California, Los Angeles*

Ratings of the likableness of persons described by 2 adjectives showed consistent violations of additive and constant-weight averaging models The effect of either adjective varied directly with the likableness of the other adjective Monotonic rescaling could remove the interactions, raising the theoretical question of whether interactions were due to nonlinearity in the rating scale or to nonadditive integration of the information Four experiments illustrate new methods for distinguishing these interpretations The fit of the subtractive model for ratings of differences in likableness between 2 adjectives supported the validity of the response scale, in addition, ratings of homogeneous combinations were linearly related to subtractive model scale values Judgments of differences in likableness between pairs of hypothetical persons, each person described by 2 adjectives, were ordinally inconsistent with additive models, confirming the interpretation that the interactions are "real" and should not be scaled away Theoretical and methodological implications are discussed

This research is concerned with how impressions of personality are formed. This topic, introduced by Asch (1946), has recently received new analytic attention (Anderson, 1968b, 1971, 1972, 1974, in press) The theoretical problem is to explain how the information provided by a set of adjectives is combined to form an overall impression In the majority of the research to be reported here, *S*'s task was to read 2 adjectives, imagine a person who would be described by both adjectives, and rate how much he would like such a person.

Figure 1 provides a schema for the analysis of impression formation The 2 stimulus adjectives are referenced by the indices, $i$ and $j$ The psychological representations of the adjectives, $s_i$ and $s_j$, are combined by the integration function, $I$, to form the psychological impression, $\Psi_{ij}$, which is transformed by the response function, $J$, to the overt response, $R_{ij}$.

There are 3 problems to be solved. (*a*) finding the appropriate stimulus representation ($s$), (*b*) determining the integration function ($I$), and (*c*) finding the

[1] This article is based on a dissertation submitted to the University of California, Los Angeles, in partial fulfillment of the requirements for the PhD degree in psychology The author was supported by a National Defense Education Act Title IV fellowship. Computing funds were received from the Campus Computing Network, University of California, Los Angeles Additional support was provided by the Center for Human Information Processing, through National Institute of Mental Health Grant MH-15828 Special thanks are due to Allen Parducci for his advice on this research and many helpful comments on earlier versions of this paper Thanks are also due to Clairice T Veit for her assistance with these experiments and to the following people for their suggestions Norman H Anderson, Bonnie G Birnbaum, Dwight Riskey, Barbara J Rose, and Chris Thomas Portions of this research were presented at the Mathematical Psychology meetings, Miami, September 1970

[2] Requests for reprints should be sent to Michael H Birnbaum, who is now at the Department of Psychology, Kansas State University, Manhattan, Kansas 66506

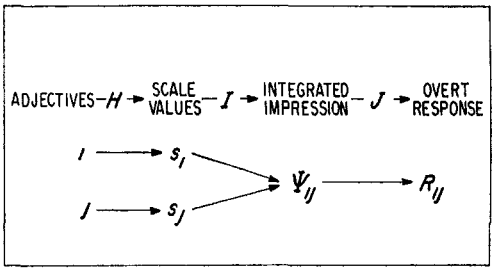


FIGURE 1 Schema for discussing research on information integration

relationship ($J$) between the overt responses and the integrated impressions Any theory of impression formation must deal, at least implicitly, with all 3 of these problems

## ADDITIVE AND CONSTANT-WEIGHT AVERAGING MODELS

A simple linear model provides an example of such a theory The adjectives are assumed to be represented as single values on a likableness continuum The $I$ function, defined on all adjective combinations, would be a linear combination of the values of the adjectives

$$\Psi_{ij} = w_0 s_0 + w_1 s_i + w_2 s_j, \qquad [1]$$

where $w_1$ and $w_2$ are the weights reflecting position in the set, $s_i$ and $s_j$ are the scale values of the adjectives, and $w_0 s_0$ is the postulated initial impression, reflecting what the impression would be in the absence of information

Anderson (1962) pointed out that if sets of adjectives were constructed according to factorial designs, then the analysis of variance test for interactions provides a powerful test of Equation 1 If the adjectives were to change their meaning nonlinearly in combination, if the $I$ function was nonadditive, or if the judgment function ($J$) was nonlinear, then significant interactions would be expected If they were nonsignificant, there would seem to be no reason to postulate such complications The data for the majority of Anderson's (1962b) 12 *S*s appeared to be in rough agreement with Equation 1

The *constant-weight averaging model* (Anderson, 1968b) is a special case of Equation 1 that requires the weights to sum to 1, the *additive* model places no restriction on the sum of the weights Critical tests of averaging vs adding models (in which the number of items in the set are varied) favor the averaging formulation over the additive (Anderson, 1971, 1974, in press) However, when the number of items in the set is held constant, as in the present experiments, the models are equivalent Therefore, the present article uses the term *additive* to include both additive and constant-weight averaging models

Anderson's (1968b) review of the evidence for the constant-weight averaging model concluded that

> The model always fits the data quite well, but there are almost always small, significant discrepancies Inspection of the data has failed to reveal the origin of the discrepancies, they may reflect some fundamental error in the model, or they may result from remaining shortcomings in the experimental technique [p 736]

Recent studies have shown large discrepancies from the constant-weight averaging model for several information integration tasks Interactive models have been proposed to account for these nonadditivities (Anderson, 1972; Birnbaum, 1972a, 1972b, 1973, Birnbaum, Parducci, & Gifford, 1971, Lampel & Anderson, 1968) However, there are 2 possible interpretations of the discrepancies that have been observed in these studies (*a*) the integration ($I$) of information is not an additive or simple averaging process, or (*b*) the judgment function ($J$) is nonlinear The next section shows how previous conceptualizations cannot differentiate these alternatives

## MODEL TESTING AND MEASUREMENT

### *Assuming Validity of Responses*

When the $J$ function is assumed to be linear, models of impression formation can be tested by comparison of theoretical predictions with the raw data. Functional measurement (Anderson, 1970) finds the

stimulus representation in accord with the model to be tested These stimulus parameters are then used as the basis for statistical tests of fit If either the model *or* the response scale was incorrect, it would show up as a significant discrepancy When the data fit the model, the fit is usually interpreted as joint support for both the model and the response scale

*Ordinal Tests*

When the $J$ function is only assumed to be strictly monotonic, it becomes more difficult to discriminate among different models Conjoint measurement analysis (Krantz & Tversky, 1971) describes conditions that ideal data would have to satisfy to be *ordinally* consistent with the theory For example, crossover interactions would be ordinally inconsistent with additive models, for no monotonic transformation could make the data fit the model. In this case, everyone agrees that the model should be rejected It should be noted that ordinal violations of additivity will show up as significant interactions in the analysis of variance The problem arises when significant interactions occur in the *absence* of ordinal violations.

*Assuming Validity of the Model*

When the additive model is assumed to be valid, then the analysis of variance tests the linearity of the $J$ function A nonlinear $J$ function will produce significant interactions even though the underlying integration is additive In the absence of ordinal violations of the model, it is possible to find a monotonic transformation, $J^{-1}$, which rescales the data to additivity

*When Both Model and Response Scale are in Doubt*

Krantz, Luce, Suppes, and Tversky (1971, p 445) have taken the extreme view that when discrepancies from additivity can be removed by monotonic transformation, they should be attributed to nonlinearity in the response scale, rather than to nonadditivity of the integration function. However, when a monotonic transformation would bring otherwise contradictory data into line with the model, the status of the model remains uncertain. This procedure *assumes* the validity of the model; hence, the existence of such a transformation does not mean that the model is validated. To resolve this difficulty, it is necessary to place transformations within the scope of psychological theory and to provide additional constraints which determine their appropriate application

## Advances in Model Testing

These 4 experiments provide a progressive sequence that systematically eliminates the additive or constant-weight averaging models The experiments illustrate novel approaches to model testing that remove the difficulty of deciding whether or not to rescale the data Since these techniques will be of interest to psychologists in many areas, they are outlined briefly below

The first 2 experiments apply criteria of stimulus and response scale consistency. Experiment I assumed the validity of a priori values for the adjectives obtained in previous work (Anderson, 1968a) and required that the integration model yield scale values that are linearly related. Experiment II investigated the effects of different response procedures, thought to produce different $J$ functions. By a principle of convergent operationism (Garner, Hake, & Eriksen, 1956), if similar interactions are obtained with different response procedures (operational definitions of $\Psi$), then the interpretation that the interactions are "perceptual" is enhanced

Experiment III illustrates how the simultaneous evaluation of 2 or more integration processes can provide the leverage to define the concept of an *appropriate* transformation (Birnbaum, 1972b, Birnbaum & Veit, 1974) *Stimulus scale invariance* requires that the scale values ($s$ in Figure 1) be independent of task *Response scale invariance* requires that the $J$ function be independent of task.

In Experiment III, *S*s performed 2 in-

tegration tasks, rating the difference in likableness between the 2 adjectives as well as the likableness of a person described by the combination Since the same stimuli and response scale were employed for both tasks, response scale invariance requires that the *same* $J$-inverse transformation be applied to both tasks The stimulus scale convergence criterion defines an *appropriate* rescaling of the combination ratings as one which *both* makes an hypothesized model fit the data *and* leads to the derivation of scale values that agree with those derived from the difference task

Although rescaling of the data might make the model fit a single set of data, the transformation that reduces the interactions implies which psychological differences are greater than or equal to others Experiment IV obtained direct ratings of these differences, and it tested whether these ratings are qualitatively consistent with the transformation that makes the data additive

Experiment IV required only the ordinal information in the data to reject the additive and constant-weight averaging models This leverage was provided by a compound integration task in which *S*s rated the difference in likableness between pairs of hypothetical persons, each described by a pair of adjectives The ratings can be rescaled to fit the subtractive model, this rescaling may or may not make the additive model of impression formation fit Thus, difference judgments can be used as a basis for response rescaling to provide a scale-free test of the constant-weight averaging model of impression formation

## Experiment I Test of Additive Models

The first experiment was designed to uncover the nature of the discrepancies previously observed Although the basic procedures were similar to those of the research described by Anderson (1968b), these were modified to permit a clearer assessment of the expected interactions Thus, only 2 factors were employed, but each was represented by a greater number of levels covering a wider range This permitted a greater variation of within-set range

### *Method*

The *S*s were presented with pairs of personality-trait adjectives and were instructed "your task is to imagine a person who would be described by *both* of the traits and judge how much you would like such a person" The *S*s recorded their judgments in numerical form, using 1 of 9 ratings for each pair 1 = dislike very very much, 2 = dislike very much, 3 = dislike, 4 = dislike slightly, 5 = neutral (neither like nor dislike), 6 = like slightly, 7 = like, 8 = like very much, and 9 = like very very much

*Subjects* The *S*s were 300 University of California, Los Angeles, undergraduates fulfilling a requirement in introductory psychology One hundred different *S*s served in each replicate of the experiment

*Stimuli* The adjectives were taken from Anderson's (1968a) list of 555 common personality-trait adjectives Each stimulus replicate consisted of a set of 25 adjective pairs produced from a $5 \times 5$ (Adjective A $\times$ Adjective B) factorial design The 5 levels of likableness of each adjective factor were separated by steps of approximately 1 28 on Anderson's 0–6 normative scale The adjectives are printed (with normative values in parentheses) in the margins of Figure 2

*Procedure* The 25 adjective pairs were printed in 1 of 6 random orders on the same page with the instructions and the labeled response scale The adjectives in each pair appeared side by side, with the adjective from Factor A on the left for half of the forms The *S*s were instructed to read through the entire list before beginning to record their ratings

### *Results and Discussion*

*Additive and constant-weight averaging models* The 3 panels of Figure 2 show the mean judgment of each adjective pair, averaged over all *S*s within the replicate The mean judgments are plotted as a function of the normative value for the adjective from Factor A, with a separate curve for each adjective from Factor B The slope of the curves represents the effects of the Factor A adjective The vertical differences between the curves represent the effects of the Factor B adjective According to additive or constant-weight averaging models, the curves in each panel should be parallel; that is, the effect of one adjective should not

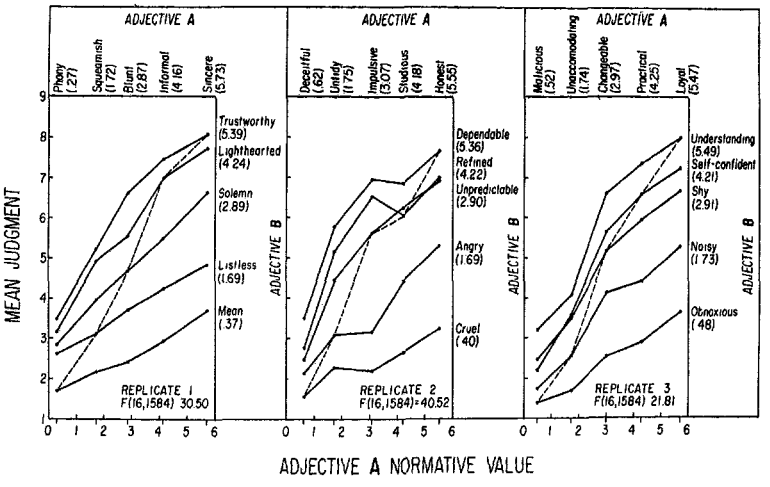


FIGURE 2 Mean ratings of likableness of pairs of adjectives Experiment I (Anderson's, 1968a, normative values for individual adjectives are listed above and to the right of each panel )

depend upon the particular adjective with which it is paired. Instead, the curves in each panel diverge to the right, the effect of either adjective appears to be proportional to the likableness of the other adjective describing the same person

This divergence was characteristic of the data for most of the individual *S*s 88% 88%, and 89% of the *S*s in the respective replicates showed this form of interaction In Experiments I–IV between 85–95% of the single *S*s resembled the group data. The A × B interactions are highly significant for each of the replicates; *F* values for these interactions are indicated in the panels of Figure 2 Of course, the *F* is much larger when the data from all 3 replicates are combined, $F$ (16, 4752) = 72 80 If the response scale of this experiment is assumed to be valid, the interactions shown in Figure 2 can be interpreted as evidence against the additivity of impressions.

Although the A × B interaction varies between stimulus replicates, $F$ (32, 4752) = 8.28, this higher order interaction is small in comparison with the similarity of the nonadditivities. There are, however, a few peculiarities that appear to depend upon the meanings of the particular adjectives For example, SELF-CONFIDENT & MALICIOUS is less likable than SHY & MALICIOUS although SELF-CONFIDENT is more likable than SHY in combination with other traits A SELF-CONFIDENT, malicious person may be perceived as more likely to carry out malicious actions than a SHY one But the divergent interaction occurs in all 3 replicates and represents the main source of difficulty for the additive models

*Multiplicative model* The multiplicative model, $\Psi_{ij} = s_i s_j$, predicts that the curves in each panel should be a diverging fan of linear functions. The interaction should be located entirely in the bilinear component (Anderson, 1970) Although the major portion of the variance of the interaction (81 6%, 56 2%, and 86 9% for the 3 replicates) was located in the bilinear component, the residuals were also statistically significant, $F$s (15, 1485) = 6 36, 20 68, and 3 26 The nature of the discrepancy can be seen most clearly in the second panel of Figure 2 According to the multiplicative model, the curves should be linear; instead, the upper curves are negatively accelerated relative to the lower curves. The other 2 replicates show the same discrepancy, but to a lesser degree.

*Range model.* The range model was fit to the data (averaged over replications) following the general procedures outlined in Birnbaum et al (1971), and using

Anderson's (1968a) normative values as estimates of the scale values The range model, using just one additional parameter, accounts for some 80% of the variance left unaccounted for by the constant-weight averaging model

*Averaging model with differential weights* The mean absolute discrepancy was only 05, indicating a very good fit to 15 points using 10 parameters The best-fit estimates of scale values for the 5 levels were 1 49, 3 02, 5 24, 6 48, and 7 83; the best-fit weights were 1 92, 1 18, 67, 73, and 90 Fewer parameters are required with the assumption of the validity of Anderson's (1968a) values and by estimating the weights as a polynomial function of scale value. Although this procedure uses more parameters than the range model, the fit was not as good

*Rescaling.* Interactions in the analysis of variance might be due to a nonlinear relationship between the impressions of likableness and the overt responses of the *S*s Thus, the actual impressions may be additive, but the overt responses may be related only ordinally to the theoretically correct response scale The possibility of response scaling was investigated using 2 techniques of data transformation

The first technique assumes that Anderson's (1968a) normative values for the single items are appropriate estimates of the scale values and that judgments of pairs of items of equal value should be linearly related to the judgments of the single items (Birnbaum, 1972a) The transformation is then estimated as a polynomial by a least squares criterion The advantage of this technique is that although it is capable of producing a radical rescaling, it will not eliminate "real" interactions that depend upon differences in within-set range, if the a priori scales are appropriate As can be seen in Figure 2, the judgments of pairs of minimal within-set range (connected by dashed lines) are already nearly linearly related to normative ratings of single values If anything, this function (treated in this procedure as the *J* function of Figure 1) is slightly negatively accelerated, so that the rescaling that makes it linear would actually increase the magnitude of the divergent interaction

If the divergent interaction is real, then the distribution of psychological impressions is somewhat positively skewed (as given by the projection of the data points on the ordinate of each panel of Figure 2) Positive skewing would lead to a negatively accelerated judgment function according to range–frequency theory (Parducci & Perrett, 1971) This suggests that contextual effects operating on the *J* function would tend to counteract the effects of true interactions Birnbaum et al (1971, Experiment V), demonstrated that manipulation of the frequency distribution of physical means of sets of psychophysical stimuli influences the form of the interaction between the components in a manner predictable from range–frequency principles of judgment

The second transformation procedure assumes that the integration is additive (or constant-weight averaging) and attempts to transform the data to fit the additive model (Kruskal, 1965) The monotone analysis of variance (MONANOVA) computer program (Kruskal & Carmone, 1969), applied to the mean judgments, greatly reduced the percentage of total variance in the interaction (from 5 0% to 4%) Thus, the data appear roughly consistent with the constant-weight averaging or additive model at an ordinal level The fact that the data can be transformed to fit raises a difficult theoretical problem Is the nonparallelism in Figure 2 due to nonadditive integration of information or to a nonlinear response function?

If the additive model is assumed to be correct, the MONANOVA analysis indicates that the judgments are a positively accelerated function of the impressions Additionally, it would mean that the scale values for adjectives presented in pairs are a negatively accelerated function of Anderson's (1968a) values for the same adjectives presented singly. The positively accelerated function for *J* derived from the MONANOVA analysis can be interpreted by range–frequency theory (Parducci & Perrett, 1971) to indicate that

the psychological distribution of stimuli is negatively skewed Kanouse and Hanson (1972) have also hypothesized that the distribution of evaluative stimuli is negatively skewed, but based their arguments on other considerations.

In short, Experiment I demonstrates that ratings of likableness are inconsistent with the additive models However, 2 interpretations are consistent with the data. The first assumes that the ratings are valid measures of psychological impressions and concludes that the integration of information is nonadditive The second assumes the validity of the constant-weight averaging (or additive) model and concludes that the responses are a positively accelerated function of the impressions Both interpretations can account equally well for the data of Experiment I Consequently, the following experiments were designed to discriminate them

## Experiment II Variation of Response Scales

The earlier experiments yielding data more consistent with the additive models have used other procedures for obtaining *S*'s response (see Anderson, 1974, in press) The interactions observed in Experiment I may be due to nonlinearity in the *J* function that depends upon the particular procedure *S* uses to indicate his impressions. Therefore, Experiment II tested the additive models using several different procedures for responding to assess whether the interaction obtained in Experiment I is specific to the 9-category rating scale used in that experiment Each of the 4 conditions of the present experiment tested a different interpretation of how the interactions might depend upon the method for responding: reversing the scale, endpoint anchoring with 20 categories, linemark responses, and matched pairs

Although the matched pairs procedure has not been used before, it has the apparent advantage that it does not require a metric response from the *S* Thus, this procedure avoids the argument that metric responding may induce *S*s to integrate separate implicit responses to the individual adjectives rather than forming an overall impression before responding

### *Method*

As in Experiment I, *S*s read pairs of personality-trait adjectives and judged how much they would like hypothetical persons who would be described by both adjectives The conditions differed with respect to the instructions for the response Each set of instructions was used in a separate small experiment, with different *S*s

For the reversed scale, 1 = like very very much and 9 = dislike very very much The *S*s were consequently instructed to rate how much they would *dislike* the hypothetical persons The 20 *S*s rated the adjective pairs of both Replicates 1 and 2 of Experiment I

The 20-category scale was anchored by instructions that 1 represented the likableness of a person who would be described by MEAN, PHONY, MALICIOUS, OBNOXIOUS, and LIAR, 20 represented the likableness of SINCERE, LOYAL, INTELLIGENT, UNDERSTANDING, and DEPENDABLE. The *S*s were instructed to judge each adjective pair relative to these end anchors and to assign a numeral between 1 and 20 to represent the appropriate position relative to the end values The 34 *S*s rated the 25 adjective pairs of Replicate 2

For the linemark response, *S*s were instructed to indicate each judgment of likableness by making a short vertical mark on a line so that the length between the margin and the mark would be proportional to the likableness The 46 *S*s in this condition judged the adjective pair of Replicate 2

The list for the matched pairs procedure was constructed from Anderson's (1968a) normative data Each pair contained 2 adjectives of equal scale value The 22 pairs covered the baseline range 35–5 60, in 25-category steps The 50 *S*s were instructed to judge each of the adjective pairs of Replicate 2 by selecting the pair of adjectives from the list of 22 pairs, "most nearly equal in likeableness to the pair you are judging" The value of *S*'s response was taken to be the normative value of the adjective pair selected by *S* The 22 pairs are listed in Birnbaum (1972b)

*Supplementary scaling* These 22 adjective pairs and the 30 adjectives used in Experiment I were printed in random order Each of 100 *S*s judged 3 aspects of each adjective or pair (*a*) the *likableness* of a person who would be described by the adjective or pair, (*b*) the *activity* of such a person, and (*c*) the *range* of likableness of persons who would be described by the adjective(s) The *S*s judged all of the adjectives on one aspect before proceeding to the next task Nine-point scales were used for all 3 scaling tasks, with 9 = like very very much, very very active, or very very wide range of possibilities

### *Results and Discussion*

Figure 3 shows the mean responses, averaged across *S*s, in each condition.

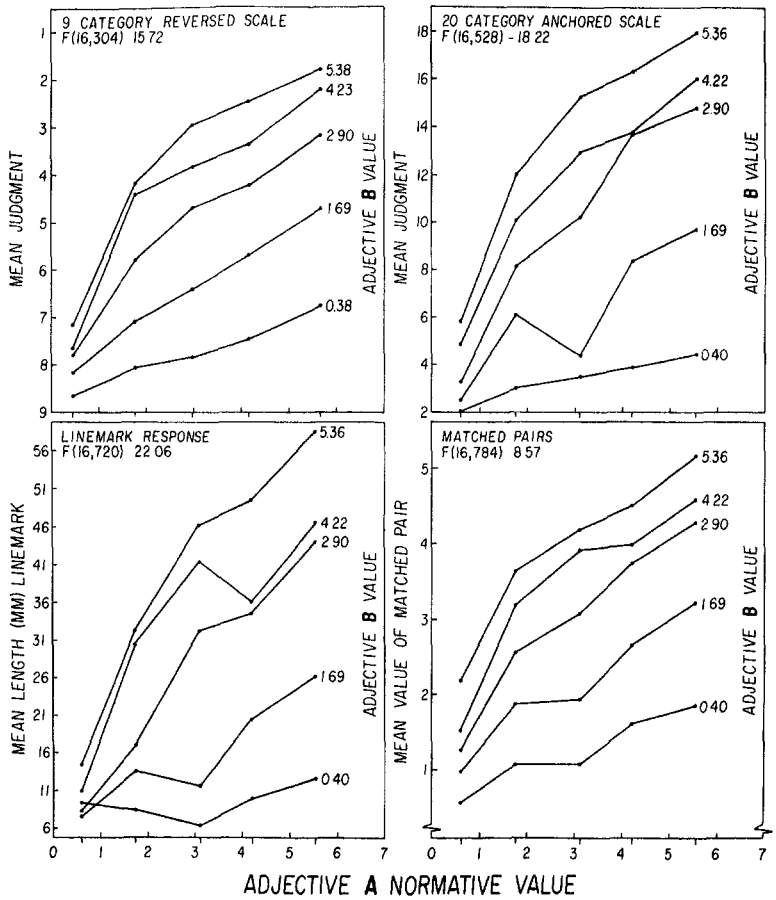


FIGURE 3 Mean judgments of likableness of pairs of adjectives Experiment II (Each panel shows results for a different response procedure )

Additive and constant-weight averaging models again predict parallel curves for each panel Instead, the interactions are all of the same general form (divergence) as in Figure 2, and the $F$ values listed in each panel show that they are all highly significant

The ordinate of the upper left-hand panel of Figure 3 has been relabeled for direct comparison with Figure 2. The data show the same interactions, and ratings of "dislikableness" appear to be a linear function (with negative slope) of ratings of "likableness " Since reversing the scale did not reverse the interaction, the deviation from additivity cannot be attributed to anything like preference for smaller numbers The interaction for the anchored scale (upper right-hand panel) shows the same divergent form as those shown in the other panels and is highly significant The percentage of variance associated with the interaction is actually larger for the linemarks condition than for any of the other conditions The positively accelerated relationship between linemarks and category ratings may account for the larger interaction.

The matching procedure assumes that the normative values obtained by Anderson (1968a) are valid estimates of scale values One check on this technique is to see whether the pairs having minimal within-set range are linearly related to Anderson's scale values As can be seen by connecting these points (the diagonal) in the lower right-hand panel of Figure 3, this assumption is supported. Never-

theless, the interaction has the same divergent form as the others and is highly significant. The results for this nonnumerical response technique do not support the objection that perhaps *S*'s integration strategy is affected by the numerical response procedures typically used in this type of study The error terms for this method were greater than for any of the other methods, perhaps because *S*s must judge the pairs of items defining the scale in addition to the usual stimulus pairs

In summary, the main conditions of Experiment II show that variation of experimental procedures for responding did not eliminate the divergent interaction. Each procedure can be considered an alternative operational definition of the impression. By the principle of converging operations (Garner et al , 1956), the divergent interactions in Figure 3 support the interpretation that *impressions* of personality are nonadditive In order to retain the additive model, it would be necessary to assume that the responses for all of these procedures are nonlinearly related to the impressions

*Supplementary scaling* One purpose of the supplementary scaling was to check a previous prediction (Birnbaum, 1972a) that items of lower value have narrower dispersions Successive interval scale values and dispersions (Torgerson, 1958) were obtained for each adjective and adjective pair. Mean ratings were linearly related to Anderson's normative values and to Thurstone scale values derived from the same data In agreement with the distributional interpretation, Thurstone scale values were positively correlated with dispersion values ($r = .65$) Since each of these dispersions was produced by the differences between the responses of different *S*s to the item, they are probably not the best estimates of the distribution of meaning for the adjectives. Only if it is assumed that different *S*s select a randomly sampled "person" from the distribution of "persons" who possess the trait would the Thurstone dispersion reflect this dispersion directly. Nevertheless, the correlation between value and dispersion seems consistent with this prediction.

Subjective estimates of the range of likableness seem a more direct measure of dispersion. These were highly correlated with mean ratings of likableness ($r = 91$) The plot of this relationship showed that although range is a nearly linear function of likableness for the dislikable traits, neutral and positive traits are all seen as implying similar wide ranges of likablenesses This finding is consistent with the fact that the interactions in Figure 2 appear to be located in combinations that include dislikable traits. In a post hoc analysis of the $2 \times 2$ subdesign containing only positive adjectives for Experiment I, the interaction was nonsignificant ($F < 1$). The $2 \times 2$ subdesign containing the negative traits had a significant interaction, $F\ (1, 297) = 5\ 93$.

In summary, the supplementary scaling suggests that adjectives should be represented by distributions, with the lower valued items having smaller variance

*Supplementary test of activity.* One interpretation of the crossover interaction observed in Experiment I, Replicate 3 (Figure 2) is that the activity component of the meaning of one adjective can multiply the evaluative component of the adjective with which it is paired. Hence, a SELF-CONFIDENT MALICIOUS person is more actively malicious than a SHY one. Similarly, an IMPULSIVE CRUEL person may be more likely to act cruel than an UNTIDY one (Experiment I, Replicate 2)

This hypothesis—that one adjective can behave like an adverb—was further investigated by having 130 additional *S*s rate the likableness of 36 hypothetical persons, each described by 2 adjectives produced from a $6 \times 6$ factorial design. The adjectives for the first factor varied in likableness (RUDE, IMMATURE, TROUBLED, DIRECT, INTELLECTUAL, and HONEST) The adjectives for the second factor were relatively neutral in likableness, but 3 were "active" (IMPULSIVE, AGGRESSIVE, and CHANGEABLE) and 3 were "passive" (QUIET, HESITANT, and SHY).

Consistent with the prediction, the likableness effect of an adjective was greater when paired with an active adjective. The interaction was large and significant,

$F$ (25, 3225) = 14 49, and even showed reliable crossovers, for example, HESITANT & HONEST is less likable than AGGRESSIVE & HONEST although HESITANT & RUDE is more likable than AGGRESSIVE & RUDE

Ratings of the activity of the 30 adjectives used in Experiment I were uncorrelated with ratings of likableness of the same adjectives but were correlated with the Thurstone dispersions ($r$ = 43) The activity hypothesis is not sufficient to account for the overall divergent interactions in Figures 2 and 3, but it appears to explain some of the second-order effects (which would otherwise be called "peculiarities") for particular adjective combinations in Experiment I.

## Experiment III Subtractive Preference versus Additive Combination

Experiment III had the following objectives· (*a*) to evaluate a subtractive model for preference judgments reflecting the *difference* in likableness between 2 hypothetical persons, each described by one adjective; (*b*) assuming that the subtractive model were to fit, to compare scale values for the preference task with those obtained for the usual *combination* task, in which *S*s judge the likableness of a person possessing both traits; (*c*) to use the scale values obtained for the preference task as the basis for rescaling the data from the combination task, (*d*) to use the scale values for the preference task to evaluate Anderson's normative values for single items; and (*e*) to evaluate the effects of a change in instructions for the combination task, in which the adjectives are given equal importance and accuracy.

The subtractive model asserts that preference ratings can be represented as the algebraic differences between the values of the items·

$$\Psi_{ij}^{D} = s_i - s_j, \qquad [2]$$

where $\Psi_{ij}^{D}$ is the psychological difference in likableness (preference for Stimulus $i$ over $j$), and $s_i$ and $s_j$ are the scale values of these 2 stimuli If the subtractive model can be fit to the data, then the adjectives can be located as points on a unidimensional likableness continuum

Two principles of scale convergence can be applied to this experiment (*a*) Response scale invariance With the same set of stimuli and the same response procedure, the judgment function is assumed to be independent of task (*b*) Stimulus scale invariance· With the same set of stimuli, the scale values of the stimuli are assumed to be independent of task. Principle *a* implies that whatever transformation is applied to fit the data to the additive model for the combination task should also be applied to the data for the difference task. Principle *b* implies that irrespective of whatever transformations of the responses fit the data for the 2 tasks to their respective models, the scale values derived from the subtractive model should be linearly related to the scale values for the same adjectives in combinations Stimulus scale invariance can be considered a necessary (but not sufficient) condition for establishing meaningful scales and psychological laws A failure of scale convergence can provide the leverage for rejecting either the subtractive or additive model, even though the $J$ functions are unknown

### *Method*

*Stimuli* The stimuli were the adjective pairs of Experiment I

*Subjects* The *S*s were 90 University of California, Los Angeles, undergraduates, 30 serving in each replicate

*Procedure* Each *S* performed both tasks, with half of the *S*s in each replicate performing the difference task first There were no effects of task order

In the difference task, *S*s were instructed to imagine 2 different people, each described by one of the adjectives of each pair, and then to judge the difference in likableness between the 2 persons The 9-point scale had labels varying from 1 = like the person (described by the adjective) on the left very very much more than the person (described by the adjective) on the right, through 5 = like both persons equally, to 9 = like the person on the right very very much more than the person on the left.

The procedure for the combination task was in all respects identical to that of Experiment I, except that the instructions were modified to emphasize that the adjectives should be considered

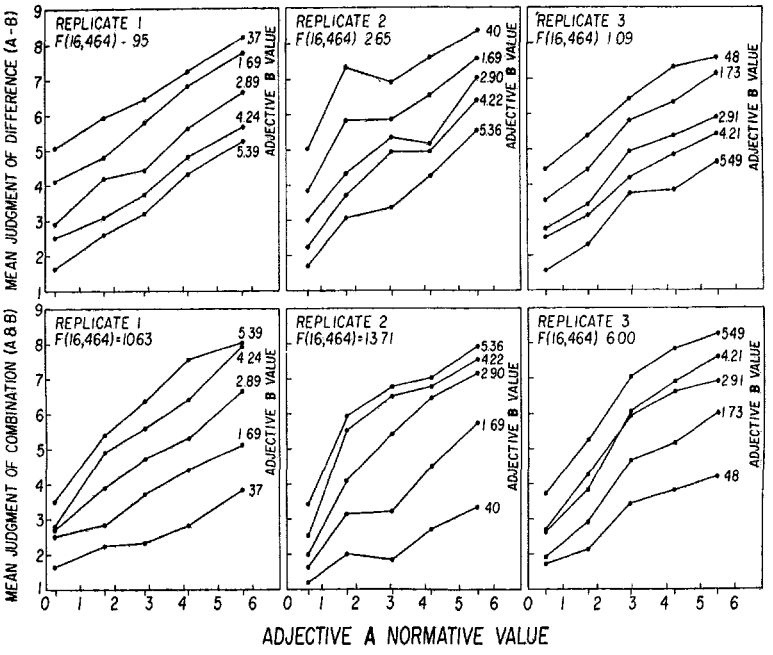


FIGURE 4 Mean ratings of difference in likableness between 2 hypothetical persons, each described by a single adjective (upper panels), and mean ratings of likableness of hypothetical persons, each described by the same 2 adjectives (lower panels). Experiment III

of equal importance and accuracy, each word having been contributed by a different acquaintance "who knows the person well"

*Results and Discussion*

*Difference task* The upper panels of Figure 4 show the mean judgments of the difference in likableness Because ratings reflect the difference between Factors A and B, the slopes are positive, but the order of the curves is negative.

Since the subtractive model is a special case of an additive model, the parallelism of the curves supports its validity. Only the interaction for Replicate 2 is significant, and examination of the figure indicates that the interaction is minor and mainly due to one point. Since Anderson's (1968a) normative values are plotted on the abscissa, the near linearity of these functions supports the validity of the normative values as estimates of scale values. Similarly, the vertical separations between the curves are nearly proportional to differences in Anderson's scale values. The good fit of the subtractive model supports the validity of the category rating scale and indicates that the adjectives can be located on a unidimensional likableness continuum.

*Combination task* The lower panels of Figure 4 show the mean judgments of the likableness of persons described by both adjectives The data appear to replicate Experiment I closely, including the peculiarities for ratings of particular adjective combinations The interactions for this experiment diverge in the same fashion as Experiment I and are highly significant for each replicate, as indicated by the $F$ values printed in the panels of the figure. The $F$ values are smaller than those for Experiment I because there were fewer $S$s in this experiment The change in instructions, emphasizing equal importance and accuracy, appears to have had no discernible effect on the interactions

*Scale convergence* Both the stimulus and response scale invariance assumptions would require that either the subtractive

model for differences or the constant-weight averaging model for combinations must be rejected Response scale invariance implies that any transformation of the ordinate of the lower panels must be applied to the upper panels as well Because the subtractive model fit the data directly, any nonlinear transformation would induce an interaction. Therefore, if the validity of the subtractive model and the principle of response scale invariance are assumed, the constant-weight averaging model must be rejected

The fit of the subtractive model implies that the marginal means for these data constitute an interval scale for the adjectives (Anderson, 1970) However, when the combination data were rescaled to additivity using Kruskal's (1965) MONANOVA, scale values derived from this procedure were nonlinearly related to those for the same adjectives derived from the subtractive model Therefore, the principle of stimulus scale invariance implies that this transformation would *not* be appropriate

Further support for the validity of the rating scale is provided by the finding that ratings of homogeneous combinations (adjectives of similar normative value) are nearly linearly related (slightly sigmoidal) to subtractive model scale values. Hence, the assumption of stimulus scale invariance implies that the $J$ function is roughly independent of task Thus, 3 procedures agree—Anderson's (1968a) normative values, subtractive model scale values, and ratings of homogeneous combinations are all nearly linearly related

This analysis also indicates that a simple multiplicative model, $\Psi_{ij} = s_i s_j$, would be inappropriate for the combination task, it would incorrectly predict that judgments of adjectives of equal value are a positively accelerated quadratic function of the scale values for the difference model However, a geometric averaging model (square root of the product of the scale values) would still be consistent with the subtractive model scale values

The data of Experiment III indicate that with the same stimuli, *S*s, and general experimental procedure, ratings of preference are consistent with a subtractive model, but ratings of combinations are again inconsistent with the additive models. This suggests that the interactions are not due to trivial details of experimental procedure If scale values for the adjectives are assumed to be independent of task, the fact that the ratings of homogeneous combinations are a nearly linear function of the scale values for the difference ratings implies that it would be inappropriate to rescale the ratings of combinations In order to retain the additive model, it would be necessary either to reject the subtractive model (in spite of its good fit to the data in Figure 4), *or* assume that *both* the scale values *and* the response function depend upon the task.

## Experiment IV Qualitative Nonadditivity

Experiment IV was designed to provide a test of additivity that would require only the ordinal information in the data. The evidence against the additive and constant-weight averaging models obtained in the first 3 experiments is illustrated in Figure 5A, where the difference in rating between $a_2b_2$ and $a_2b_1$ is greater than the difference between $a_1b_2$ and $a_1b_1$ An additive model requires that these differences be subjectively equal If the $J$ function is assumed to be linear, then differences in judgment are proportional to differences in the impressions, therefore, the divergence (previously shown in Figures 2, 3, and 4) would be contrary to additive models However, if $J$ were positively accelerated, impressions might be additive The same ratings have been transformed to parallelism in Figure 5B Although the difference in rating (Figure 5A) due to the change from $b_2$ to $b_1$ depends on whether $a_1$ or $a_2$ was in the same set, the difference in psychological value (Figure 5B) is assumed by additive models to be constant.

Experiment IV tested this possibility by asking *S*s to judge directly the differences in likableness between integrated impressions. The success of the subtractive

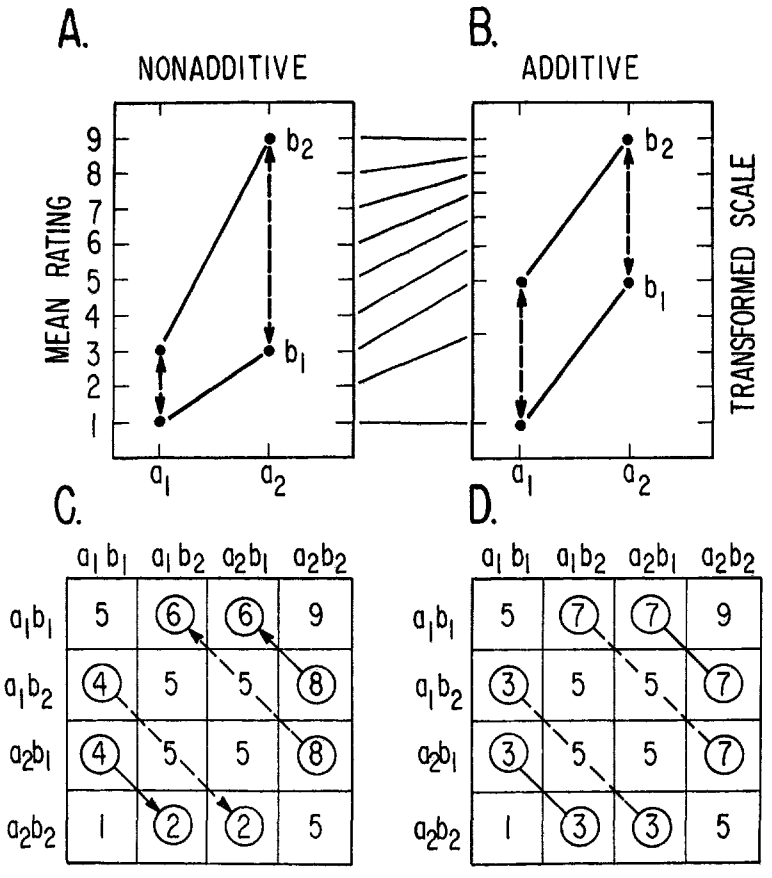


FIGURE 5 A Schematic diagram of nonadditive data B Schematic diagram of additive impressions (following transformation) showing prediction for difference judgments (note that differences in transformed ratings are assumed to represent psychological differences) C Matrix of hypothetical data representing ratings of the difference in likableness between the person described by the pair of adjectives for the column and the person described by the pair for the row, assuming nonadditive impressions but subtractive preferences D Matrix of hypothetical data for column minus row differences, assuming additive impressions and subtractive preferences [Matrix entries for Panels C and D are equal to $\frac{1}{2}(d) + 5$, where $d$ is the difference between the ordinate values for Panels A and B ]

model in Experiment III suggests that it can be used to infer the impressions of the combinations It is only necessary to assume that ratings of differences in likableness are strictly monotonic to differences in psychological impressions If the interaction is real, then the preference for $a_2b_2$ over $a_2b_1$ should exceed the preference for $a_1b_2$ over $a_1b_1$, as illustrated by the length of the dashed arrows in Figure 5A If the interaction reflects only a nonlinearity in the $J$ function, however, then the 2 preferences should be equal, as shown by the dashed arrows of Figure 5B.

The matrices in the lower panels of Figure 5 illustrate 2 patterns of data that might be obtained for ratings of the difference in likableness between 2 persons, each described by a pair of adjectives Figure 5C shows the pattern that would be obtained if the integration of information were nonadditive Each pair of circled entries is an example of a predicted rank-order comparison The dashed ar-

TABLE 1

MEAN RATINGS OF THE DIFFERENCE IN LIKABLENESS BETWEEN PAIRS OF HYPOTHETICAL PERSONS

| Person on the left | Person on the right | | | |
|---|---|---|---|---|
| | PHONY & MEAN | PHONY & TRUSTWORTHY | SINCERE & MEAN | SINCERE & TRUSTWORTHY |
| Replicate 1 | | | | |
| PHONY & MEAN | 4 91 | 6 34 | 6 31 | 7 92 |
| PHONY & TRUSTWORTHY | 3 83 | 5 07 | 4 94 | 7 42 |
| SINCERE & MEAN | 3 54 | 4 89 | 4 94 | 7 46 |
| SINCERE & TRUSTWORTHY | 1 66 | 2 60 | 2 51 | 5 01 |
| | DECEITFUL & CRUEL | DECEITFUL & DEPENDABLE | HONEST & CRUEL | HONEST & DEPENDABLE |
| Replicate 2 | | | | |
| DECEITFUL & CRUEL | 5 03 | 6 15 | 6 26 | 8 20 |
| DECEITFUL & DEPENDABLE | 3 65 | 4 92 | 4 88 | 7 89 |
| HONEST & CRUEL | 3 89 | 5 26 | 5 06 | 7 95 |
| HONEST & DEPENDABLE | 2 09 | 2 46 | 2 12 | 4 83 |
| | MALICIOUS & OBNOXIOUS | MALICIOUS & UNDERSTANDING | LOYAL & OBNOXIOUS | LOYAL & UNDERSTANDING |
| Replicate 3 | | | | |
| MALICIOUS & OBNOXIOUS | 5 03 | 6 35 | 6 37 | 8 20 |
| MALICIOUS & UNDERSTANDING | 3 77 | 5 00 | 5 25 | 7 58 |
| LOYAL & OBNOXIOUS | 3 58 | 4 85 | 5 06 | 7 62 |
| LOYAL & UNDERSTANDING | 1 98 | 2 55 | 2 40 | 5 12 |

rows represent the same comparison as the dashed arrows in Figure 5A Since the ratings represent judgments of differences between impressions labeled by the column and the row entries, the direction of the order is reversed below the diagonal Figure 5D represents the type of pattern that would be obtained if the adjectives combined additively. Each pair of circled entries would be equal Both matrices in Figure 5 were generated by assuming the subtractive model for preference (Equation 2) However, the rank order of the matrix entries would remain the same if subjected to a strictly monotonic transformation Thus, the data can be rescaled to fit the subtractive model without precluding the ordinal test for the additive integration models for impression formation (Birnbaum, 1972b)

For example, if the interaction shown in Figure 2 is real, then the judged difference in likableness between someone who is UNDERSTANDING & LOYAL and someone who is UNDERSTANDING & MALICIOUS should exceed the judged difference between someone who is OBNOXIOUS & LOYAL and someone who is OBNOXIOUS & MALICIOUS But if adjectives were integrated according to an additive or constant-weight averaging model, then the 2 differences would be equal

*Method*

The *S*s were instructed to judge the difference in likableness between pairs of hypothetical persons, each of whom was described by 2 adjectives Two adjectives were printed on the left side of the page and 2 on the right side of the page The *S*s were instructed to form impressions of the personalities of the 2 persons before judging the difference in likableness The 2 adjectives describing each person were described as being of equal accuracy and importance, *S*'s task was to imagine that they were contributed by different, but equally reliable, sources Judgments were in terms of a 9-point scale, labeled as in the difference task of Experiment III

*Subjects* The *S*s were 195 University of California, Los Angeles, undergraduates, 65 in each of 3 stimulus replicates

*Stimuli* The stimulus for each judgment consisted of 4 adjectives, 2 printed on the left side of the page and 2 on the right The adjective pairs describing the "person on the left" were produced from a $3 \times 3$, A $\times$ B, factorial design, the adjective pairs on the right were constructed from a $2 \times 2$, C $\times$ D, factorial design Each person on the left was combined with each person on the right producing a $3 \times 3 \times 2 \times 2$, (A $\times$ B) $\times$ (C $\times$ D), factorial design

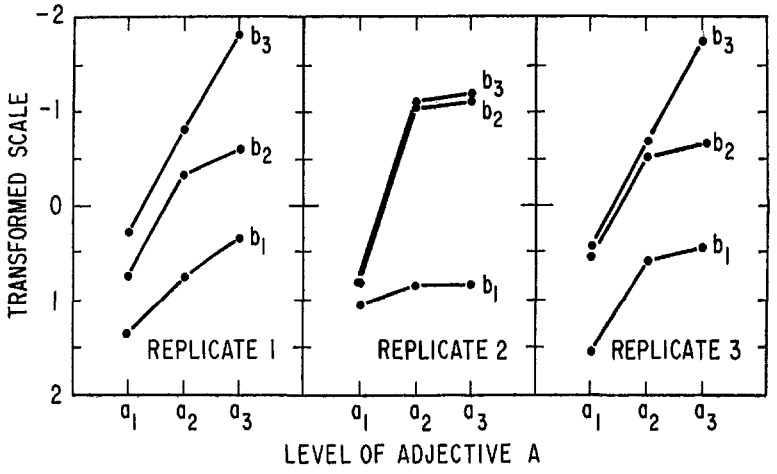


FIGURE 6 Transformed impression values for the person on the left derived from MONANOVA applied to difference ratings Experiment IV (The 3 levels of subscripts refer to low, medium, and high likableness adjectives of Experiment I For example, $a_1$ = PHONY, $a_2$ = BLUNT, and $a_3$ = SINCERE, for Replicate 1 To assess metric information in mean ratings, compare this figure with Figure 2 )

The basic idea of the stimulus construction can be understood by consideration of Figure 2 For the third replicate, the adjectives were $a_1$ = MALICIOUS, $a_2$ = CHANGEABLE, $a_3$ = LOYAL, $b_1$ = OBNOXIOUS, $b_2$ = SHY, $b_3$ = UNDERSTANDING, $c_1$ = MALICIOUS, $c_2$ = LOYAL, $d_1$ = OBNOXIOUS, and $d_2$ = UNDERSTANDING With respect to Figure 5C, the last entry in the first row (9) would represent the rating of the difference in likableness between someone who is LOYAL & UNDERSTANDING and someone who is MALICIOUS & OBNOXIOUS Nine additional cells in the design were constructed for each replicate by pairing each person on the left with an additional person on the right who was INFORMAL & LIGHTHEARTED Therefore, the entire design represents a 9 × 5 factorial design with the factors representing the person on the left and the person on the right (see Birnbaum, 1972b)

*Procedure* The 45 trials were printed in 1 of 3 different random orders in booklets with different page orderings for different *S*s The *S*s read through the entire list before beginning to record their judgments

## *Results and Discussion*

The mean judgments of difference in likableness are presented in Table 1 for that part of the design illustrated in Figure 5 Comparison of the rank order of the ratings with the direction of the differences for the circled values shows that the pattern resembles that of Figure 5C representing an interaction. Differences in rating from Experiments I, II, and III predict the ratings of differences in the present experiment

The crucial comparisons (circled entries in Figure 5D), which must be equal for the additive or constant-weight averaging model, are significantly in the direction predicted by the interactions in Figures 2–4 for all 3 replicates, *F*s (1, 64) = 71 13, 115.88, and 89 54, respectively. The individual ratings by single *S*s show the same pattern of rank orders for these crucial comparisons as the mean judgments in the table Of the 195 *S*s, 81 were in perfect ordinal agreement with the interaction obtained in the first 3 experiments for all 4 comparisons. Only 12 *S*s showed a greater number of rankings in the direction opposite from that predicted by the interaction, compared with 170 *S*s whose rankings were in the direction predicted by the interactions obtained in Experiment I. These results are clearly inconsistent with the additive model which would predict an equal split of these rankings between the 2 directions.

The larger, 9 × 5 design was used to test the subtractive model for preferences between combinations. The subtractive

model also predicts parallelism, and the data appeared roughly parallel when plotted However, statistically significant interactions were obtained for each stimulus replicate, $F$s (32, 2048) = 3 72, 4 98, and 5 10 for Replicates 1, 2, and 3, respectively Application of MONANOVA to these ratings of preference reduces the interaction to less than 1% of the total variance, so it is apparent that the preference ratings are at least ordinally consistent with the subtractive model

Following the application of MONANOVA to the 9 × 5 subtractive model design, the derived impression values for the combinations were inconsistent with the additive and constant-weight averaging models Figure 6 plots the impression values derived from the MONANOVA transformation for the 9 persons on the left Since these were produced from a 3 × 3 factorial design of adjectives, the nonparallelism of these scale-free values refutes the additive models of impression formation. Similar results were obtained for the 2 × 2 design for persons on the right

In summary, the minimal assumption that the preference ratings are monotonically related to subjective differences leads to the conclusion that the additive and constant-weight averaging models must be rejected Comparison of scale-free values in Figure 6 with the scale-dependent ratings in Figures 2, 3, and 4 shows that ratings contain metric information that correctly describes the divergent interaction

Krantz et al (1971, pp 445–447) have argued that if an interaction can be removed by monotonic transformation, then it should be attributed to the rating scale and should not be given a psychological interpretation However, Experiment IV shows that this procedure could lead to erroneous conclusions Although the interaction obtained in Experiment I could be removed by monotonic transformation, ratings of differences in impression are predictable from the differences in ratings This indicates that the interactions are of psychological significance Rescaling those data to parallelism would be inappropriate, as demonstrated by the qualitative contradictions of Experiment IV

However, the ratings also appear to contain some nonlinearity that reduces the apparent magnitude of the interaction for nonnegative items It would be improper reasoning to conclude that nonnegative items combine additively based on the lack of interaction in the ratings The scale-free values in Figure 6 appear to show a continued divergence Nonlinearity in the response scale may actually make nonadditive data appear additive The procedures of Experiment IV avoid the rescaling problems and provide a truly scale-free test of the additive models

## General Discussion

This research demonstrates that impressions of likableness cannot be represented as simple sums or constant-weight averages of the values of the adjectives Instead, one bad trait results in an unfavorable overall impression with the other trait having less influence This effect reflects a real psychological interaction and cannot be attributed to the response scale

Some of the early research on impression formation concluded that the constant-weight averaging model could give a reasonable fit to ratings of adjective combinations The parallelism prediction appeared to be roughly satisfied, and this finding was interpreted to validate both the model and the rating scale Systematic deviations did occur in this research, but were often attributed to experimental difficulties or nonlinearity of the response scale However, it is now clear that the constant-weight averaging model is not an appropriate general description, but may give a reasonable approximation when the stimulus range is restricted and the experimental design lacks power [3]

[3] An alternative view might maintain that the constant-weight model is descriptive of impression formation, but only under limited experimental circumstances The $E$ would be advised to carefully select adjectives, instructions, response procedures, and other conditions to minimize the nonparallelism When the data appear parallel, this view contends that $E$ would have the right to conclude that the parallelism jointly supports the constant-weight model and the response scale However, this approach seems unsatisfactory for several reasons First, the present research shows that different response procedures, "equal impor-

*Theoretical Implications*

It is useful to consider a set of conditions that yield the parallelism prediction to consider what nonparallelism might mean. The following conditions underlie the parallelism prediction of the averaging model (*a*) the integration function is an averaging process; (*b*) the adjectives within each factor have equal absolute weight, (*c*) the adjectives do not change value nonlinearly in combination, nor do the weights depend upon the particular stimulus configuration, and (*d*) the *J* function is linear Based on a single experiment such as Experiment I, nonparallelism could be interpreted as evidence to disprove at least one of the premises Without further constraints, it would be impossible to specify whether nonparallelism was due to a nonadditive *I* process (Conditions *a–c*) or a nonlinear *J* function (Condition *d*) These experiments provide the leverage to indicate that the nonparallelism is *not* attributable to the *J* function. There are several remaining possibilities The integration function may not be an averaging process, the stimulus parameters may depend upon the configuration, or the weights may not be equal for the adjectives

The differential-weight averaging model (Anderson, 1971, 1972, 1974) can account for the interactions by allowing the weight of an item to depend upon its scale value Differential weighting requires many more parameters and seems unnecessarily complicated to fit the simple divergence interactions The model can account for a wide variety of interactions, making it difficult to disprove. However, with large enough designs, there are degrees of freedom left over to permit tests (e g , Anderson, 1972; Birnbaum, 1973) Configural weighting (Birnbaum, 1972b), in which the weight of an item depends in part on its rank within the set, requires fewer parameters and gives a slightly better fit to the data.[4]

The configural-weight model predicts steady divergence for Figures 2, 3, 4, and 6. The differential-weight model is more flexible and could account for reconvergence, among other patterns It is thus of theoretical interest to ask if the curves show any evidence of reconvergence The scale-dependent ratings of Figure 2 appear nearly parallel for nonnegative traits However, the scale-free values in Figure 6 show steady divergence for Replicates 1 and 3 The ratings may contain a small scale-end effect that reduces the apparent interaction The bulge in Figure 2, Replicate 2 is apparently due to IMPULSIVE multiplying the effect of the other adjective, rather than having less weight on its own Since there is no evidence for more than a simple divergence, the existing data cannot test between differential and configural-weight models Methods for distinguishing these models have been suggested by Birnbaum (1973)

*Methodological Implications*

Previous conceptualizations of the stimulus concatenation problem have not separated the

tance and accuracy" instructions, and different-selections of adjectives all yield similar divergent interactions If the model is to be deemed correct, then it must apply under very special conditions indeed Second, since it is possible to manipulate the form of *J*, it follows that selection of experimental procedures could lead to an "experimental rescaling" of the data Thus, it should be possible to select end anchors and filler stimuli to reduce the interaction in the analysis of variance Therefore, it would be inappropriate to select experimental procedures that yield parallel data and then conclude that the parallelism is joint support for model and response scale That parallelism could result from a combination of nonadditive *I* and nonlinear *J* is not merely an untestable philosophical possibility but represents a plausible hypothesis that can be tested by the procedures of Experiment IV.

[4] An averaging mechanism can be analogized to a balance plank and fulcrum In the differential-weight model, each adjective corresponds to a weight ($w$) placed at a certain location ($s$) on the plank The integrated impression ($\Psi$) is the location where the fulcrum must be placed so that the plank will balance This location is the weighted average, $\Psi = \sum w_i s_i / \sum w_i$ The configural-weight averaging model assumes that the weight of a stimulus depends upon its rank within the set to be judged For 2 stimuli, the simple range model (Birnbaum et al , 1971), $\Psi_{ij} = .5(s_i + s_j) + \omega|s_i - s_j|$, can be rewritten as a configural-weight model, $\Psi_{ij} = (5 + \omega)s_i + (5 - \omega)s_j$, when $s_i > s_j$ This model can also be represented by a balance plank mechanism, however, the configural-weight model does not require each location on the plank to have its own weight For 2 stimuli in the set, there would be only 2 weights, one for the lower valued item, and one for the higher valued item. The same weight can be placed at any location on the plank, if it holds the same rank This model becomes a *minimum* model when $\omega = -5$, a *constant-weight* model when $\omega = 0$, and a *maximum* model when $\omega = 5$, and can describe a family of simple convergent or divergent interactions.

$J$ and $I$ functions of Figure 1. This article offers a conceptualization that separates these problems. The propriety of rescaling the data has been uncertain in previous work. Functional and conjoint measurement (Anderson, 1970; Krantz et al., 1971) both allow for monotone transformation, but differ in their outlook about when transformation is appropriate.

Krantz et al. (1971, pp. 445–447) reanalyzed the data of Sidowski and Anderson (1967) and concluded that since the interaction could be eliminated by a monotonic transformation of the mean ratings, the original authors were incorrect in attributing psychological significance to their findings. An investigator following this rescaling procedure with the data of Experiment I would have erroneously concluded that the additive model was an appropriate description of impression formation. The present research, by demonstrating that data transformation can lead to erroneous conclusions, provides a warning against this practice.

Anderson[5] originally attempted to rescale the data for ratings of the severity of disturbed behaviors and to attribute the nonadditivity to the rating scale, but he has recently reinterpreted the same data as evidence for the configurality of clinical judgment by fitting these data to an averaging model with differential weights (Anderson, 1972). The latter interpretation assumed the linearity of the $J$ function.

The present research supports this attitude toward rating scales, but it also provides scale-free constraints that make the interpretations of Krantz et al. (1971) and Anderson (1972, see, also, Footnote 5) matters for experimentation rather than assumption.

In Experiment IV it would be possible to distinguish multiplicative from additive models on the basis of ordinal information (see Figure 5). The simple view of conjoint scaling would not allow this distinction, since a monotonic (logarithmic) transformation of a positive product yields a sum. By assuming that ratings of differences are monotonically related to subjective differences, the $I$ function can be separated from the $J$ function. Experiment IV allows rescaling, but rescaling does not preclude the possibility of testing the $I$ function.

Although the rating scale contains metric information, it does not seem appropriate to assume that $J$ is linear in every experiment. The $J$ function can be predictably nonlinear. It depends upon contextual effects that are explainable (Birnbaum et al., 1971; Parducci & Perrett, 1971). A nonlinear $J$ function could result in 2 possible errors: (*a*) nonadditive data when the underlying impressions are additive, or (*b*) additive data when the impressions are nonadditive. Thus, the decision to rescale or not to rescale cannot be convincingly settled without further constraints, such as those applied in Experiments III and IV.

[5] N. H. Anderson. Application of an additive model to impression formation. Paper presented at the third annual meeting of the Psychonomic Society, St. Louis, August 1962.

## *Conclusions*

Impressions of likableness cannot be represented as simple sums or averages of single values of the adjectives. They appear to be a predictable, but nonadditive function of the component values. The data show consistent, regular deviations from additivity that are similar for different selections of adjectives. When one adjective is dislikable, the person is rated as dislikable, and variation of the other trait has less effect.

Differential or configural weighting of the more dislikable traits can account for the interactions. Representation of the stimuli by distributions could explain why the adjectives are integrated by a nonadditive function. It is less likely for a person possessing a dislikable trait to be likable than for a person with a likable trait to be dislikable.

Finally, this research illustrates that rating scales contain metric information that should not be uncritically scaled away. The procedures employed in Experiments III and IV provide constraints that allow a model to be tested without having to assume the validity of the rating scale.

## REFERENCES

ANDERSON, N. H. Application of an additive model to impression formation. *Science*, 1962, **138**, 817–818.

ANDERSON, N. H. Likableness ratings of 555 personality-trait words. *Journal of Personality and Social Psychology*, 1968, **9**, 272–279. (a)

ANDERSON, N. H. A simple model for information integration. In R. P. Abelson, E. Aronson, W. J. McGuire, T. M. Newcomb, M. J. Rosenberg, & P. H. Tannenbaum (Eds.), *Theories of cognitive consistency: A sourcebook*. Chicago: Rand McNally, 1968. (b)

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

ANDERSON, N H. Integration theory and attitude change. *Psychological Review*, 1971, **78**, 171–206

ANDERSON, N H Looking for configurality in clinical judgment *Psychological Bulletin*, 1972, **78**, 93–102

ANDERSON, N H Cognitive algebra Integration theory applied to social attribution In L Berkowitz (Ed ), *Advances in experimental social psychology* Vol 7 New York Academic Press, 1974

ANDERSON, N H Information integration theory A brief survey In D H Krantz, R C Atkinson, R D. Luce, & P Suppes (Eds ), *Contemporary developments in mathematical psychology* Vol. 2 New York· Academic Press, in press.

ASCH, S E Forming impressions of personality *Journal of Abnormal and Social Psychology*, 1946, **41**, 258–290

BIRNBAUM, M H Morality judgments Tests of an averaging model. *Journal of Experimental Psychology*, 1972, **93**, 35–42 (a)

BIRNBAUM, M H The nonadditivity of impressions Unpublished doctoral dissertation, University of California, Los Angeles, 1972 (b)

BIRNBAUM, M H Morality judgment Tests of an averaging model with differential weights *Journal of Experimental Psychology*, 1973, **99**, 395–399

BIRNBAUM, M H , PARDUCCI, A , & GIFFORD, R K Contextual effects in information integration *Journal of Experimental Psychology*, 1971, **88**, 158–170

BIRNBAUM, M H , & VEIT, C T Scale convergence as a criterion for rescaling Information integration with difference, ratio, and averaging tasks. *Perception & Psychophysics*, 1974, **15**, 1–9

GARNER, W R , HAKE, H W , & ERIKSEN, C W Operationism and the concept of perception. *Psychological Review*, 1956, **63**, 149–159

KANOUSE, D E , & HANSON, L R , JR *Negativity in evaluations* Morristown, N J . General Learning Press, 1972.

KRANTZ, D H , LUCE, R D., SUPPES, P., & TVERSKY, A *Foundations of measurement* New York Academic Press, 1971

KRANTZ, D H , & TVERSKY, A Conjoint measurement analysis of composition rules in psychology *Psychological Review*, 1971, **78**, 151–169

KRUSKAL, J B Analysis of factorial experiments by estimating monotone transformations of the data *Journal of the Royal Statistical Society* (Series B), 1965, **27**, 251–263

KRUSKAL, J B , & CARMONE, F J MONANOVA· A FORTRAN-IV program for monotone analysis of variance *Behavioral Science*, 1969, **14**, 165–166

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 , & PERRETT, L F Category rating scales Effects of relative spacing and frequency of stimulus values *Journal of Experimental Psychology Monograph*, 1971, **89**, 427–452

SIDOWSKI, J B , & ANDERSON, N H Judgments of city-occupation combinations *Psychonomic Science*, 1967, **7**, 279–280

TORGERSON, W S *Theory and methods of scaling* New York Wiley, 1958

(Received November 16, 1972)