

AVERAGING VERSUS ADDING AS A STIMULUS-COMBINATION RULE IN IMPRESSION FORMATION

NORMAN H. ANDERSON

University of California, San Diego

Ss rated their liking of persons described by sets of 2 or 4 personality-trait adjectives. Adjectives were chosen from 4 scale ranges: extremely or moderately favorable or unfavorable. Results were as follows. (a) A set of 2 moderate and 2 extreme adjectives produced a less extreme response than the set of 2 extreme adjectives alone. This result is inconsistent with the idea that S reaches his impression by adding the stimulus values; it is qualitatively consistent with the idea that he averages the stimulus values. (b) With all adjectives of equal value, an increase in the number of adjectives per set produced a more extreme response. It was shown how this set-size effect, which seems to be inconsistent with an averaging model, can be handled by such a formulation. (c) The additive and averaging models made the same predictions for 2 quantitative comparisons of which 1 showed significant discrepancy. There is thus some question whether either formulation can handle the data at a quantitative level.

Suppose you considered "pains-taking" to be a moderately desirable trait, and "well-spoken" to be highly desirable. Would you then like a "pains-taking, well-spoken" person more than a "well-spoken" person?

The answer depends on how you integrate the stimulus information. The simplest theoretical analysis would start with the assumption that each stimulus has a value, and that the overall evaluation is obtained by combining these values by some simple rule. Two obvious combination rules are that the stimulus values are added, or else averaged, with the resultant sum or mean of the stimuli given as the response. The answer to the question above would then be Yes if you added, and No if you averaged the stimulus values.

Among those who have studied tasks of this kind, some have favored

¹ This work was done at the University of California, Los Angeles, and was supported by National Science Foundation Grant GB-1470. I wish to thank Ann Jacobson for her assistance.

an averaging-type formulation (e.g., Anderson, 1959, 1962a, 1965; Levy & Richter, 1963; Osgood, Suci, & Tannenbaum, 1957; Rimoldi, 1956; Spence & Guilford, 1933; Weiss, 1963; Willis, 1960), and others have favored an additive formulation (e.g., Abelson, 1961; Gulliksen, 1956; Hammond, 1955; Johnson, 1955; Triandis & Fishbein, 1963).

Neither formulation may be entirely adequate, but there is not much quantitative evidence. Much of the work has been limited to showing some measure of agreement between predicted and observed. However, almost any model will produce substantial correlations between predicted and observed if a wide range of stimulus values is used. Moreover, high correlations can be obtained even when there are large discrepancies from prediction. Those few studies that have used a reasonably adequate test of goodness of fit have yielded conflicting results. Both the positive and negative findings

require closer scrutiny than can be given here, but the evidence suggests that at least one of the formulations has some validity.

For the most part, the cited studies have used stimulus sets of equal size. It then requires special measures to be able to distinguish between additive and averaging models. Indeed, there is as yet no serious evidence comparing the two formulations except perhaps for the Spence and Guilford (1933) study of odor combinations. The primary purpose of the present experiment was to get a relative test of the additive and averaging formulations based on qualitative comparisons.

Existing evidence indicates that an increase in the size of the set of stimuli produces a more extreme response (Anderson, 1959; Podell, 1962; Stewart, 1965; Willis, 1960). This result might itself appear to favor an additive formulation and to rule out an averaging model on qualitative grounds. From previous work, however, it was clear that this result could be handled at a qualitative level within an averaging formulation. The second purpose of the present experiment was to attempt a more quantitative analysis of the set-size effect. Finally, the experimental design provided an incidental test of the quantitative accuracy of both the additive and averaging models.

METHOD

In the experimental situation, *S* read a set of personality-trait adjectives that described a person, and rated how much he himself would like a person so described. The instructions stated that all adjectives of a set were equally important and should be given equal attention. Sets of two or four adjectives were used, and the main variable was the scale value of the adjectives.

Stimuli.—The adjectives were taken from a master list of 555 common personality-trait

adjectives. This list had been previously rated by 100 *Ss* on a 7-point, Dislike-Like scale that ranged from 0 to 6.² Four sublists of 32 adjectives each were used: *H*, from 5.00 (reasonable) to 5.45 (truthful); *M*⁺, from 3.45 (painstaking) to 3.74 (persuasive); *M*⁻, from 2.22 (unpopular) to 2.54 (dependent); and *L*, from 0.72 (spiteful) to 1.00 (abusive).

There were 10 types of set, corresponding to the following combinations of adjectives from the four scale ranges listed above. Combinations of two adjectives from each separate scale range yielded four set types, and these are denoted as *HH*, *M*⁺*M*⁺, *M*⁻*M*⁻, and *LL*. Similarly, combinations of four adjectives from each separate scale range yielded four more set types: *HHHH*, *M*⁺*M*⁺*M*⁺*M*⁺, *M*⁻*M*⁻*M*⁻*M*⁻ and *LLLL*. The remaining two types of set also had four adjectives corresponding to the patterns *HHM*⁺*M*⁺ and *LLM*⁻*M*⁻.

The sets of adjectives were constructed to use a fairly large sample of adjectives, to avoid bias in set construction, and to allow comparisons based on the same adjectives where logically possible. The construction details are not necessary to an understanding of the experiment.

Each stimulus replication consisted of 32 sets obtained as follows. First, two pairs of adjectives were randomly chosen from each scale range, yielding 8 sets of two adjectives each. The two pairs in each scale range were also combined to form 4 sets of four adjectives each. Finally, 4 more sets of four adjectives were formed by randomly pairing the 2 *HH* sets with one or the other of the *M*⁺*M*⁺ sets, and the 2 *LL* sets with one or the other of the *M*⁻*M*⁻ sets. This procedure yielded 16 sets and it was repeated with new adjectives for a total of 32 sets.

Four such stimulus replications were constructed and included as a factor in the experimental design. Adjectives were chosen without replacement so that each replication was based on different adjectives and all 32 adjectives in each scale range were used.

In addition to the 32 experimental sets,

² A copy of this list has been deposited with the American Documentation Institute. Order Document No. 8503 from ADI Auxiliary Publications Project, Photoduplication Project, Photoduplication Service, Library of Congress, Washington, D. C. 20540. Remit in advance \$1.25 for microfilm or \$1.25 for photocopies and make checks payable to: Chief, Photoduplication Service, Library of Congress.

there were 10 practice sets, 1 each of the 10 set types. Each adjective that was used in the experimental sets for a given *S* appeared exactly once in the practice sets for that *S*.

Procedure.—Each *S* received two mimeographed booklets, one of practice sets, and one of experimental sets. Except for the cover sheet, each page of the booklet contained two adjectives on each of one or two lines, together with a blank space in which *S* recorded his response. For HHM⁺M⁺ sets and LLM⁻M⁻ sets, the adjectives appeared in alternating order which was balanced over the two sets of each type. The pages of each booklet were arranged in a different scrambled order for each *S*.

The *S* was instructed to read the adjectives slowly to himself. A total of 15 sec. was used for each set. At the end of the allotted time, *E* said, "Rate & Turn." The *S* then recorded his response and turned to the next page.

The *Ss* were told to use 50 to rate a person they would neither like nor dislike, to use lower numbers for persons they would dislike, and to use higher numbers for persons they would like. Except for centering the scale at 50, *Ss* were told they could use any numbers they wished. It was emphasized that they should feel free to change their use of the rating scale at any time during the practice sets.

Subjects.—The *Ss* were 48 males who were fulfilling a course requirement in introductory psychology. They were run one or two at a time, and were randomly assigned in equal numbers to the four stimulus replications.

RESULTS

Table 1 lists mean response per set for each of the 10 set types. There are three main aspects of the data that are of interest.

The primary purpose of the experiment was to get a comparative test of the additive and averaging formulations in the way alluded to in the first paragraph of the introduction. The critical comparison is based on the HH and the HHM⁺M⁺ sets. Both sets have the same highly favorable adjectives, HH; the latter set has also the added moderately favorable adjectives, M⁺M⁺. The addition of this moderately favorable information should increase the favor-

TABLE 1
MEAN RESPONSE PER SET AS A FUNCTION OF SET TYPE

Set Type	Response	Set Type	Response
HH	72.85	HHHH	79.39
M ⁺ M ⁺	57.56	M ⁺ M ⁺ M ⁺ M ⁺	63.20
M ⁻ M ⁻	42.18	M ⁻ M ⁻ M ⁻ M ⁻	39.50
LL	23.70	LLLL	17.64
		HHM ⁺ M ⁺	71.11
		LLM ⁻ M ⁻	25.67

Note.—H, M⁺, M⁻, and L stand for Highly Favorable, Moderately Favorable, Moderately Unfavorable, and Highly Unfavorable adjectives. Response scale is centered at 50; higher numbers indicate favorable impressions, lower numbers indicate unfavorable impressions.

ableness of the response if the stimuli are combined by an additive process. Thus, an additive model would predict that the response to HHM⁺M⁺ would be higher than the response to HH alone. The observed difference is in the opposite direction: The mean response of 71.11 for HHM⁺M⁺ is less than the mean response of 72.85 for HH.

The LL and LLM⁻M⁻ sets give a similar comparison. Both sets have the same highly unfavorable LL adjectives, but the latter set has also the added moderately unfavorable M⁻M⁻ adjectives. An additive model would thus predict a lower response to LLM⁻M⁻ than to LL. The observed difference is again in the direction opposite to prediction: The mean response of 25.67 for LLM⁻M⁻ is higher than the mean response of 23.70 for LL.

To test these results, each *S* was given a difference score for each comparison, the direction of difference being taken so that an additive model would predict a positive score. The overall mean difference of -1.86 was significantly less than zero, $F(1, 44) = 6.50$. The stimulus-replication factor did not approach significance, indicating that the results did not arise from a fortuitous choice.

adjectives. It is thus concluded that, for the present stimuli at least, the addition of moderately polarized information to highly polarized information decreases the polarity of the response. This conclusion, which contradicts a simple additive model, is consistent with an averaging formulation.

Since the critical result is qualitative, resting on a direction of difference, it cannot reasonably be attributed to a shortcoming in the scale of measurement. Moreover, adding moderately polarized to moderately polarized adjectives actually increases the polarity of the response. This can be seen in the comparisons between sets of two and four M⁺ or M⁻ adjectives (Table 1, second and third lines). This set-size effect supports the interpretation of the critical result in two respects. First, it validates the nominal neutral point of the rating scale and the choices of moderately polarized adjectives. Second, it underscores the contrary direction of the effect in the critical comparison.

The second aspect of the results is the effect of set size. These data bear on the adequacy of an averaging formulation. In agreement with previous findings cited above, Table 1 shows that for adjectives of equivalent value, four adjectives produced a more polarized response than two adjectives. The mean magnitude of the difference, averaged over the four scale ranges, is 5.23; the corresponding mean standard error is 1.02.

This result would seem contrary to an averaging formulation. If the response equaled the mean scale value of the stimuli, it would be the same for sets of two and sets of four adjectives. However, the data may be accounted for by making use of an initial impression, analogous to the

initial opinion employed in previous work with an averaging-type model for serial presentation (Anderson, 1959). It is assumed that S has an initial impression, I_0 , which is combined in a weighted average with the scale values of the stimuli of each set. The response to a set of Size 1 is then given by

$$R_1 = wA + (1 - w)I_0, \quad [1]$$

where A is the scale value of the single item, w is the weight associated with the item, and $1 - w$ the weight for the initial impression.

For sets of Size k in which all stimuli have the same weight and scale value, the equation for the response is

$$R_k = \frac{k w A + (1 - w) I_0}{k w + (1 - w)}. \quad [2]$$

In this expression, the numerator is the weighted sum of the scale values, and the denominator is just the sum of the weights.

To apply these equations, the HH and LL sets were considered as of Size 1, with scale values of 100 and 0. I_0 was set equal to 50, the nominal neutral point of the response scale. Equation 1, together with each S 's observed mean response to the HH or LL sets, was then used to estimate w for that S . Equation 2, with $k = 2$, was then applied to predict S 's response to the sets of four adjectives. The mean predicted response to HHHH was 79.7 which is .3 higher than the obtained value; the mean predicted response to LLLL was 16.6 which is 1.0 lower than the obtained value. (These predictions differ from those obtainable from the group data of Table 1 because the mean ratio does not equal the ratio of the means.)

Although the agreement between predicted and obtained is fairly good,

