COMBINING VISUAL AND VERBAL INFORMATION IN AN IMPRESSION-FORMATION TASK

ANITA K. LAMPEL AND NORMAN H. ANDERSON

Stanford University University of California, San Diego

Female students rated males characterized by a photograph and 2 personality-trait adjectives in terms of desirability as a date. The experimental design tested how these 3 component pieces of information were integrated to form the overall impression. The results supported the hypothesis that the 2 adjectives were simply averaged together, but there was a strong interaction between adjectives and photograph. The adjectives had less effect in the context of the less desirable photographs. This was interpreted as a discounting effect, in line with the social nature of the judgment task. It was noted that the averaging model would be consistent with the interaction if discounting produced lower values of the weight parameters of the adjectives. The averaging formulation was also consistent with the comparisons between ratings of photographs alone and photographs plus adjectives.

Most experiments on impression formation use homogeneous sets of stimuli so that the subject bases each impression upon several items from the same stimulus class. The most-used task is that introduced by Asch (1946), in which the subject gives his impression of a person described by several personality-trait adjectives. A few investigations have employed other homogeneous sets of stimuli such as photographs (e.g., Levy & Richter, 1963; Willis, 1960), or meal dishes or life events (Anderson & Norman, 1964).

The present experiment, in contrast, uses heterogeneous sets of stimuli. In the experimental task, female subjects judged how much they would like to date a male characterized by two personality-trait adjectives and a photograph. To obtain her impression, the subject must then integrate two distinct types of stimulus information, verbal and visual.

Relatively little work has been done with heterogeneous stimuli. However, the report of Kerrick (1959) is particularly relevant here, both experimentally and theoretically. Kerrick presented combinations of ambiguous newspaper photographs and verbal captions which her subjects rated on a good-bad scale. The results contradicted the congruity formulation of Osgood and Tannenbaum (1955) and were interpreted in favor of a summation formulation.

The present experiment was designed to get further information on how subjects integrate heterogeneous information. In addition, it incorporated a comparison between heterogeneous and homogeneous information. Since each judgment is based on two adjectives and a photograph, there are two theoretical problems to consider.

The first problem concerns the way in which the subject integrates the two homogeneous pieces of verbal information. Previous work with sets of adjectives (e.g., Anderson, 1962, 1968) has tended to support a weighted average model. In the model, each stimulus item is characterized by two parameters, a scale value and a weight. The first parameter reflects the meaning of the stimulus item with respect to the judgment scale employed (likableness, etc.); the second parameter represents the relative importance of the stimulus item in the overall impression. The judgment is then taken to be the sum of the scale values each multiplied by its respective weight.

The present task differs from previous work in that the integration of the verbal information is done in the context of disparate relevant information. However, the experimental
design allowed a test of the model for the adjective combinations within that context. The second problem concerns the way in which the subject integrates the verbal with the visual information. One possibility is that the photograph is effectively treated as verbal information. In that case, the averaging model could be applied directly to these heterogeneous combinations. The experimental design allowed a test of this hypothesis also.

The other possibility that was initially considered was that some kind of multiplicative formulation might apply. Cliff (1959) has employed such a model for adverb-adjective combinations. For the present stimuli, the adjectives might be considered to modify or intensify the photographs. Accordingly, a multiplicative model might be appropriate for the data.

METHOD

In the experimental task, the subject judged word-photo combinations on desirability as a date. Each combination consisted of a photograph of a male student together with two adjectives that were taken to be characteristic of his personality. The subject rated how much she would like to date that student on a 20-point scale ranging from 1, "I would very much dislike dating this person," to 20, "I would very much like dating this person."

Procedure. Each word-photo combination was presented for 10 seconds after which the experimenter said, "Rate," and the subject responded. After the practice series, the instructions were reviewed and the test series presented for the first time. Following a short break, the test series was presented a second time, the subject being told that natural fluctuations in judgments were to be expected. Finally, the photos were presented alone and rated in the same way.

The word-photo combinations were presented on cards as follows. Each photo was 1X1.25-inch in size and was mounted at the top of a 4X6 card. Each adjective pair was typed on a 1.50X3-inch card. To present a given combination, the specified adjective card was placed on the specified photo card. To present a given combination, the specified adjective card was placed on the specified photo card and the set was then placed in front of the subject. The word-photo combinations were presented in one of two randomized orders in the test series.

Stimuli. The visual stimuli were six photographs of male Caucasian students. In preliminary work, two of these photos had been classified as "very acceptable as a date," two as neutral, and two as "not acceptable as a date."

The verbal stimuli were personality-trait adjectives selected from a standardized list (Anderson, 1965, Footnote 2). Two different adjective replications were chosen and are shown in Table 1. Each replication consists of four pairs of adjectives, of varied normative scale value as listed in the left column of the Table. It should be noted that the four pairs of adjectives in each replication form an 2X2 design.

The experimental test series consisted of 24 word-photo combinations. These were obtained by pairing each of the 6 photos with each of the 4 adjective pairs in one adjective replication.

In addition, there were seven practice combinations, as well as six other combinations interspersed in the test series. These combinations used the same photos combined either with medium-value adjectives, or else with extremely high or extremely low adjectives to serve as end-anchors for the response scale.

Instructions and subjects. There were two instructional conditions, "equal accuracy" and "naturalistic." The equal accuracy instructions told the subject that both adjectives were equally accurate and important. The naturalistic instructions were similar to those used by Anderson and Jacobson (1965); they suggested that the two adjectives in a combination might not be equally reliable or important.

The subjects were 40 Caucasian females, single and under 21, who were fulfilling a course requirement in introductory psychology. Ten were randomly assigned to each of the four conditions obtained from two sets of instructions and two adjective replications.

Statistical analysis. The primary tests of the model are based on single-subject analyses of variance. The basic element of the overall experimental design is a 6X2X2 factorial composed of six photographs and the four pairs of adjectives from the 2X2 design. Since each subject judged each of these 24 combinations twice, for each single-subject analysis df = 24 for error. A number of pooled analyses were also made as noted below. Instructions and adjective replications were not significant and will not be discussed further unless specifically noted.

RESULTS

The mean judgments of the combinations are shown in the first four columns of Table 2. The data are pooled over the two adjective replications and over the two photographs at each level. The increase in response down each column reflects the strong effect of the value of the photo on the judgment.

<table>
<thead>
<tr>
<th>Type</th>
<th>Replication 1</th>
<th>Replication 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>HiM</td>
<td>friendly gracious</td>
<td>appreciative restless</td>
</tr>
<tr>
<td>HLM</td>
<td>friendly self-contented</td>
<td>showy showy</td>
</tr>
<tr>
<td>LHM</td>
<td>messy gracious</td>
<td>appreciative</td>
</tr>
<tr>
<td>LM</td>
<td>messy self-contented</td>
<td>showy</td>
</tr>
</tbody>
</table>

* H, M, and L denote traits of high, medium, and low normative value.
These data bear on three questions about the way in which the stimulus information is integrated. The first question is how the two adjectives of a given pair are combined in the judgment. The hypothesis that they combine in a simple linear manner implies that the adjective-adjective interaction of the 2 x 2 design of Table 1 is zero (Anderson, 1962, 1968). In terms of the data, the test is whether the difference between the first two column means is equal to the difference between the second two column means. The algebraic size of this interaction is .4 scale point which does not approach significance in the pooled data, F(1,36) = 2.48. However, in the single-subject analyses, 8 of the 40 subjects (20%) had interactions that were significant at the .05 level. Only 2 of these 8 were in the equal accuracy group, though the difference between the groups was not significant. On the whole, it would seem that at least a minority of the subjects did have real adjective-adjective interactions.

The second question is how the verbal and visual information are combined. The hypothesis that they combine in a linear manner implies that the word-photo interaction is zero. For this hypothesis, the outcome is clear. The pooled data show highly significant F ratios of 28.96 and 19.13, df = 5/180 for each, for the interaction of photographs with the two factors of the adjective design. More directly, 27 of the 40 individual subjects, about equally divided between the two instruction conditions, showed interactions that were significant at the .05 level. Accordingly, it may be concluded that there is a moderately strong and prevalent tendency toward such verbal-visual interaction.

The nature of the word-photo interaction is such that the word has greater effect with the more desirable photos. This can be seen explicitly in Figure 1. The upper curve shows the effect of the first factor of the 2 x 2 adjective design. That is, it plots the mean difference between the effect of Traits HI and L as a function of the value of the photo. A horizontal line would mean zero interaction. The observed upward slope of the data shows that the words carry progressively more influence as the value of the photo increases. The lower curve of Figure 1 shows similarly the effect of the second factor (H2 versus M) of the adjective design. Here again the descriptive traits have greater effect when combined with the high than with the low photos. The difference between the two curves themselves is as expected. It simply reflects the greater differential effect of HI versus L as compared to H2 versus M.

Finally, it is of interest to compare the judgments of the word-photo combinations with the judgments of the photos alone as shown in the fifth column of Table 2. On the hypothesis that the values of the words

---

TABLE 2

<table>
<thead>
<tr>
<th>Photograph value</th>
<th>Adjective pair</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>H1;H2</td>
</tr>
<tr>
<td>Low</td>
<td>5.8</td>
</tr>
<tr>
<td>Medium</td>
<td>11.7</td>
</tr>
<tr>
<td>High</td>
<td>17.2</td>
</tr>
<tr>
<td>M</td>
<td>11.6</td>
</tr>
</tbody>
</table>
and photos are simply added or summated, this fifth column of data should differ from any of the other four columns by an additive constant. This is clearly not so. Indeed, if the data are taken at face value, then the second column shows that the H₁M adjective pair adds to the value of a low photograph and subtracts from the value of a high photograph. Thus the failure of a summation rule does not result from such possible difficulties as ceiling or floor effects.

The present results, therefore, do not completely agree with those of Kerrick (1959), who reports a predominant summation effect. For further detail, Dunnett’s test (Winer, 1962, p. 89) was applied to the data of each row of Table 2, with the photo-alone condition serving as control. In each row, three of the four word-photo means were significantly different from the corresponding mean for photo alone. In particular, both the H₁M and the L₁H₂ adjective pairs produce a significant increase in the response to the low photo, and a significant decrease in the response to the high photo. These combined results contradict any summative formulation, linear or nonlinear.

If only the medium photos are considered, the present data are in agreement with Kerrick’s results. Since Kerrick’s photos were rather weakly polarized, there is no disagreement between the present results and hers at the empirical level. However, the inclusion of more highly polarized photos in this experiment has led to a different theoretical outlook.

It may be worth pointing out that the foregoing comparison involves a general methodological problem which, though it does not seem serious here, could cause difficulties in other cases. In the strictest sense, the photos with and without words are two distinct classes of stimuli. In such cases, comparison of the overt responses will ordinarily rest on some assumption of a corresponding comparability of the psychological process underlying the judgments of the two classes of stimuli. Thus, it is logically possible that the subjects shifted their standard of judgment from one class of stimuli to the other. Although this would not seem to be a serious danger here, an analogous comparison of word-photo combination with words alone would perhaps require more explicit justification.

There is also the more concrete consideration that the judgment of a given stimulus will in general depend on the range and distribution of other stimuli with which it is being judged (e.g., Parducci, 1963). That is, the subject’s frame of reference and his use of the response scale will be influenced by the range and distribution of stimuli. It is not impossible that the frame of reference gradually shifted over the course of presentation of the word-photo combinations or of the subsequent photos alone. This does not seem serious here, in the first respect because the frame of reference should have stabilized during practice, and in the second respect because there were only six photos and these were presented immediately after the combinations. In other experiments, this could present a problem. One routine precaution would involve two presentations of the set of stimuli. A shift in the frame of reference would then be revealed by systematic shifts between the judgments of the first and second replications. In the present experiment, two presentations were made of the combinations, though not of the photos alone. The relevant statistics are the main effect of and (especially) the interactions involving presentations, and these were not significant.

**DISCUSSION**

For adjective-adjective combinations, the present results are in agreement with previous reports that these homogeneous stimuli combine by an averaging rule. This integration rule is not quantitatively perfect, however. As in previous work that has used adequate tests of fit, the present results show some apparently real though small discrepancies from the model. Their source has not been found, and it is possible that they reflect a serious flaw in the model. Tentatively, however, the present experiment extends previous findings, since here the adjective information was integrated concurrently with the qualitatively different visual information.

For the adjective-adjective combinations, it should be noted, the present design does not
distinguish between averaging and summating as stimulus-combination rules (Anderson, 1968). This could be done, however, by varying the number and value of the adjectives accompanying a given photograph as has been done for adjectives alone (Anderson, 1965, 1967). Such heterogeneous combinations would be particularly important in testing the model, and would be interesting in their own right as well.

For the adjective-photograph combinations, on the other hand, there is no doubt about the outcome. The size and nature of the interaction definitely rules out any simple linear model. At the same time, comparison of the judgments of photos with and without adjectives gives a direct indication that no summation model is applicable.

The alternative possibility that had originally been considered was that the adjective would act as a modifier and intensifier of the photograph. On that basis, a multiplicative formulation might be expected to apply. The upper curve of Figure 1 is, at least approximately, a fixed multiple of the lower curve. This would be consistent with a purely formal application of a multiplicative model. Substantive considerations, however, suggest that the high and low photographs have positive and negative values, respectively, with the scale zero in between. If this is so, then a multiplicative model would predict U-shaped interaction curves in Figure 1. For if the adjectives exerted a multiplicative effect, then the magnitude of that effect would presumably vary directly with the polarity of the photograph. It would be small when the photograph was near zero or neutral in value, large when the photograph had large value, positive or negative. The observed curves in Figure 1 are clearly inconsistent with such a multiplicative formulation.

The averaging formulation can, however, account for the results if a special assumption is added. In the model, the judgment is taken to be a weighted average of the values of the components. The weight for any component corresponds to its importance in the combination. The above tests of the model rely on the assumption that the adjectives have equal weight among themselves, as seems justified on the basis of previous work. Similarly, the tests assume that the photographs have equal weight among themselves. The special assumption is that the weight of the photograph varies inversely with its value or, equivalently, that the adjectives have lower relative weight in the context of the less desirable photographs.

If indeed the adjectives carry reduced weight relative to the low-value photograph, then the difference in the effect of two adjectives would be small in such combinations. For the more valued photographs, the adjectives would carry greater relative weight, and the differential effect of any two adjectives would increase. This interpretation thus agrees completely with the observed adjective-photograph interaction that is seen in Figure 1.

This averaging formulation is reasonable from a psychological viewpoint. The social nature of the task provides a rationale for the assumption of an inverse relation between the scale value and effective weight of the photograph. There are negative sanctions on dating a male with bad physical appearance, regardless of positive personality traits that he might have. On that basis, the adjectives would be largely discounted (Anderson & Jacobson, 1965) when combined with the low photograph. As physical appearance improves to neutral or positive values, increasing importance can be given to the personality traits.

An averaging formulation is also consistent with the observed relations between the judgments of photos with adjectives and of photos alone. For instance, the HiM and LHg adjective pairs are near neutral in value. Averaging the value of either pair with the value of the photo would then increase the value of the low, decrease the value of the high, and exert relatively small effects on the value of the medium photograph. Just this pattern of results is seen in Table 2.

If this interpretation is correct, then similar results may be expected under other conditions. For example, there are homogeneous combinations that would contain one trait of unique importance, love and faith in a mate, for example, or virtue in a priest. Other adjectives combined with such prime traits would then be expected to interact
with them in the same way as they did with the photographs here.

Such traits may be similar to what Asch (1946) has called central traits. Asch's interpretation, however, is apparently in terms of contextual changes in meaning or value. The present interpretation is in terms of weight, and the above remarks on the social nature of the task suggest that the subject actively discounts the adjectives combined with the low photograph.

It should be noted, however, that similar results will be obtained if it is simply assumed that the low photographs have greater natural or absolute weight than the high photographs. Since the averaging model requires the weights to sum to unity, the adjectives will then necessarily receive lower relative weight with the low photograph. Thus, although the discounting interpretation seems reasonable in the present task, the averaging model allows for the same pattern of results more generally.

REFERENCES


(Received June 9, 1967)