

How do children form impressions of persons? They average

CLYDE HENDRICK, CHRISTINE M. FRANZ, and KENNETH L. HOVING
Kent State University, Kent, Ohio 44242

The experiment reported was concerned with impression formation in children. Twelve subjects in each of Grades K, 2, 4, and 6 rated several sets of single trait words and trait pairs. The response scale consisted of a graded series of seven schematic faces which ranged from a deep frown to a happy smile. A basic question was whether children use an orderly integration rule in forming impressions of trait pairs. The answer was clear. At all grade levels a simple averaging model adequately accounted for pair ratings. A second question concerned how children resolve semantic inconsistencies. Responses to two highly inconsistent trait pairs suggested that subjects responded in the same fashion, essentially averaging the two traits in a pair. Overall, the data strongly supported an averaging model, and indicated that impression formation of children is similar to previous results obtained from adults.

Studies of impression formation in the Asch (1946) tradition have usually found that subjects average individual traits to form an overall evaluative impression of a person (Anderson, 1965; Hendrick, 1968). In fact, Anderson (1974) has shown in numerous studies that trait as well as other information is often integrated by the simple rule $R = \sum W_i S_i$, where R is the response to the stimulus compound, W_i is the weight of item i , S_i is the scale value of i , and $\sum W_i = 1$. This weighted average formula accounts for a very extensive set of social judgment data.

Most of this research has used college students as subjects. Almost no developmental research has been done. In a rare exception, Butzin and Anderson (1973) had children judge attractiveness of pairs of toys and found that an averaging model predicted the judgments quite well. In contrast, a study by Peevers and Secord (1973) which used a free response system found considerable developmental change in the direction of more sharply differentiated person conceptions with increasing age.

The question of how children integrate information in forming impressions is important for several reasons. The answer is of intrinsic interest in its own right, but the answer also has a bearing on the interpretation of the vast amount of data already collected with adult subjects. For one thing, the true nature of the underlying psychological process remains in some doubt. The fact that impression responses fit an averaging model does not mean that the underlying process is necessarily averaging (Anderson, 1974, p. 88). Adult subjects may be following some higher order rule of "splitting the difference" whenever several discrepant stimulus items are presented to them for judgment. A more subtle consideration is that adult subjects have

acquired the formal concept of averaging, and may apply this rule at some minimal level of awareness as a higher order cognitive strategy. At the microlevel, the integration process might occur in a very different manner (perhaps unverbizable) which is then summarized and translated into overt judgment by the application of a previously learned tool rule—some variant of the arithmetic average.

Developmental research with children provides possible answers for some of these issues. For example, a 5-year-old child does not possess the formal concept of an average, does not know how to compute an average, and lacks the repertoire of other higher order judgment strategies that adults may often use. If, despite these omissions, the 5-year-old does follow an averaging rule in forming impressions, strong evidence would then exist that averaging is indeed a basic psychological process—presumably at the microlevel.

Such reasoning constituted the basic rationale behind the present study. Subjects in Grades K, 2, 4, and 6 formed impressions of strangers described by either one or two traits and rated their liking for the strangers. The experiment was designed so that it could be determined: (a) whether children at different grade levels follow a definable integration rule, (b) whether such a rule might best be classified as "averaging" or something else (such as "adding"), and (c) if an averaging model is appropriate, whether a simple average or a weighted average best describes the impression judgments.

A second concern of the experiment was to determine how children handle strong semantic inconsistencies. Adults do react minimally to such inconsistency under special experimental arrangements (Anderson & Jacobson, 1965; Hendrick, 1972), but there are no data on how children resolve such inconsistencies. Two trait pairs which were highly inconsistent, as well as the individual traits forming the pairs, were rated by all subjects, allowing a preliminary assessment of how small children react to semantic inconsistencies.¹

Reprint requests should be addressed to Clyde Hendrick, Department of Psychology, Kent State University, Kent, Ohio 44242. We wish to thank the students, teachers, staff, and administration of the University School, Kent State University for their cooperation and participation.

METHOD

Subjects

Forty-eight students from the university laboratory school, 12 each from Grades K, 2, 4, and 6, served as subjects in the experiment. Sex distributions were approximately equal: five males and seven females in Grade K; six males and six females in Grade 2; five males and seven females in Grade 4; and seven males and five females in Grade 6. The average ages of the children in Grades K, 2, 4, and 6 were, respectively, 69.1, 93.3, 120.7, and 132.0 months. Subjects were tested individually by an adult female stranger in sessions averaging between 15 and 20 min. The subject was seated at a desk opposite the experimenter with a rating scale placed directly in front of him.

Rating Scale

All judgments were made on a rating scale consisting of a graded series of seven circular faces, each 5 in in diam. The faces were cut from squares of yellow felt, and the features were cut from strips of black felt. Each face was glued to a 7-in square of white posterboard. The eyes and nose were represented as circles and were equal in size and position for all faces. Mouths were varied. Three mouths had upward curving smiles and three had downward curving frowns. The mouth in the middle of the series was represented as a straight line. The smiles and frowns were graded in degree by varying the degree of curvature. The subject made a response by pointing to the appropriate face. The faces were scored numerically on a 1- to 7-point basis.

Procedure

The subject was instructed to respond to verbal descriptions of a person by pointing to the face that corresponded to how much he thought he would like or dislike such a person. If he thought he would like the person, he was to point to one of the happy faces, and so on for each of the faces. The instructions were presented in standardized form, but were elaborated if subjects showed difficulty in understanding. After the instructions, each subject was asked a few questions about the use of the scale to insure his ability to use it correctly.

Stimulus Adjectives

Twenty-two adjectives were selected which were deemed appropriate for children. Eleven were preselected as desirable adjectives (nice, glad, happy, brave, kind, friendly, smart, right, funny, wonderful, good), and the remaining 11 were deemed undesirable (angry, bad, sad, strange, silly, mean, wrong, mad, afraid, terrible, unhappy).

The subjects initially rated these 22 single traits, with the experimenter using a quasirandom order of presentation for different subjects. With the exception of silly, which had an average positive rating, all of the mean desirability ratings conformed to initial expectation.²

Based on the initial ratings of the 22 traits, the experimenter rapidly constructed 11 trait pairs (only 6 pairs are considered in this report—see Note 1) and then obtained ratings of these pairs (described below).

After the pair ratings, half of the subjects (randomly chosen) at each grade level rerated the 22 individual adjectives in order to obtain stability data. The low variance for many of the traits made ordinary correlation coefficients inappropriate. As a substitute measure, a correlation was computed for each subject, based on the subject's 22 pair ratings of the 22 traits.

The six correlations ranged from .68 to .91 for Grade K, with a median of .82. For Grade 2 the range was .87 to .94, with a median of .91. For Grade 4 the range was .84 to .95, with a median of .93. For Grade 6 the range was .77 to .97, with a median of .88. The data indicated that at all grade levels the subjects were reasonably stable in assigning their ratings to individual trait words.

Design for Pairs

The six pairs were constructed from the individual traits according to the following design: *averaging test*—two HM^+ pairs and two LM^- pairs, where H, L, M^+ , and M^- represent positive, negative, moderately positive, and moderately negative traits, respectively. Because only a 7-point scale was used, it was decided to select HM^+ pairs of values 7 and 5 primarily (and 6 and 5 secondarily), and LM^- pairs of values 1 and 3 primarily (and 2 and 3 secondarily). A clear test of an averaging hypothesis was provided in the following manner. An averaging model requires that the response to the HM^+ and LM^- pairs be significantly less polarized than the response to the H or L singletons, respectively (see Anderson, 1965). Consistent nonsignificant differences between HM^+ pairs and H singletons and LM^- pairs and L singletons would provide support for some other model, most likely some version of what is called the "adding" model.

A test of a simple averaging model was provided by comparing the HM^+ pair rating to the arithmetic average of the H and M^+ singleton ratings. A comparable comparison was provided by the LM^- pair and L and M^- singletons. Consistent nonsignificant differences would provide support for a simple averaging model. Significant deviations of the pair ratings from the arithmetic average of the singletons would provide support for either a weighted average or an adding model.

Within each pair type, one pair was presented in a more extreme to less extreme order (HM^+ and LM^-), and the remaining pair was presented in the reverse order (M^+H and M^-L). No sequence effects were evident in the data, and this variable is ignored in the presentation of the results.

Semantic inconsistency: Two sets, good-bad and kind-terrible, were used. These two sets were always presented last following the other pairs. The traits were read in the H to L direction.

RESULTS

Averaging as a Combination Rule

The relevant data for desirable sets are given in Table 1 and for undesirable sets in Table 2. The means for the highly polarized singletons (H or L) are given in column 1 for each grade level, and the means for less polarized singletons (M^+ or M^-) are given in column 2. The arithmetic averages of the two singletons are given in column 3, and the mean pair ratings are given in column 4. These means are based on both stimulus replications of each set type.

Appropriate F ratios for model tests are given in columns 5 and 6. These F tests are based on difference scores. For each subject a difference was obtained between the rating of the extreme trait in column 1 and the pair rating in column 4. These difference scores were then tested against a null hypothesis of zero (Anderson, 1960), separately for each grade level. These F tests provide information on the viability of an averaging hypothesis, and it is these F ratios that are recorded in column 5. The F ratios in column 6 follow the same logic, except that the difference scores were simply the difference between the pair rating and the arithmetic average of the singletons which composed the pairs.

The means in column 4 compared to the means in columns 1 and 2 indicate that subjects engaged in an

Table 1
Means and F Tests for Desirable Sets

Grade Level	1 H	2 M ⁺	3 Average H + M ⁺ /2	4 HM ⁺ Pair	5 F 1-4	6 F 3-4
Kindergarten	7.00	5.04	6.02	6.00	21.2*	< 1
Second	7.00	5.08	6.04	6.00	27.6*	< 1
Fourth	7.00	5.00	6.00	6.30	19.0*	3.65
Sixth	6.92	5.00	5.96	6.21	21.4*	2.65

Note—F ratios are based on difference scores. Column 5 presents the test for the difference between H and HM⁺ sets. Column 6 presents the test for the difference between the arithmetic average of H and M⁺ and the HM⁺ pairs. The relevant null hypothesis for each test is a mean difference of zero.

* $p < .01$

averaging process in making the pair ratings. Strong confirmation is provided by the F ratios in column 5. For desirable sets, in every case the HM⁺ pair was rated as significantly less desirable than the polarized H trait entering into the pair. For the undesirable sets, in every case the LM⁻ pair was rated as significantly more desirable than the polarized L trait entering into the pair. These results are strong evidence for an averaging process. Further, it is evident that averaging occurred in about the same magnitude at all grade levels. Subjects in Grade K averaged just as consistently as subjects in Grade 6.

The F ratios in column 6 indicate that (with one exception) a simple averaging model can adequately account for the data. The pair ratings did not deviate significantly from the arithmetic average of the two singletons entering into the pair. The exception occurred for Grade 4 for undesirable sets. In that instance, the L trait apparently carried more weight, resulting in a significant deviation from a simple average. Overall, however, these data strongly support a simple averaging model.

Semantic Inconsistency

The results for the two semantically inconsistent trait pairs are given in Table 3. Means for H, L, and HL pairs are given in the columns for each grade level, as well as the arithmetic average of the H and L singletons.

For these two trait pairs, initial ratings of the singletons was considered as a random variable, and a regular analysis of variance was performed on the data with the ratings of H, L, and HL pairs considered as

three levels of a within-subjects variable and grade level as a between-subjects variable. A separate analysis was performed for each replication. The results were very clear, as shown by the Newman-Keuls comparisons given in Table 3. In every case, the HL pair was rated as significantly less desirable than the H trait and significantly more desirable than the L trait composing the pair.

These results suggest that, when faced with semantic inconsistency, the children engaged in an integration process which strongly resembled the averaging results noted above. Means are not definitive, however, since such averages would obtain if every subject discounted (see Anderson & Jacobson, 1965) one of the two traits entirely, with half of the subjects discounting the H trait and the other half discounting the L trait. Such discounting did not occur. Out of a total of 96 ratings of inconsistent pairs, 70 of these ratings were the neutral "four." The results were so strong as to suggest a simple semantic rule: One good pooled with one bad equals one neutral.³

DISCUSSION

The results demonstrated quite clearly that children integrate trait information in a fashion similar to adults. A simple averaging model adequately accounted for the results for pair ratings. This fact is particularly striking when one considers that such averaging occurred at all age levels. The results provide strong support and generality for the types of models that Anderson (1974) has advocated.

Table 2
Means and F Tests for Undesirable Sets

Grade Level	1 L	2 M ⁻	3 Average L + M ⁻ /2	4 LM ⁻ Pair	5 F 1-4	6 F 3-4
Kindergarten	1.17	2.96	2.07	1.78	7.4*	1.65
Second	1.08	3.00	2.04	2.29	23.7**	1.05
Fourth	1.00	3.17	2.09	1.75	23.0**	5.43*
Sixth	1.42	3.29	2.35	2.00	5.9*	3.51

Note—Explanation for F ratio is the same as for Table 1.

* $p < .05$

** $p < .01$

Table 3
Means for Inconsistent Sets

Grade Level	Replication 1				Replication 2			
	H Good	L Bad	HL Pair	Average H + L/2	H Kind	L Terrible	HL Pair	Average H + L/2
Kindergarten	5.92 ^a	2.58 ^b	3.58 ^c	4.25	5.92 ^a	1.92 ^b	3.83 ^c	3.92
Second	5.75 ^a	2.50 ^b	3.67 ^c	4.13	6.17 ^a	1.50 ^b	3.50 ^c	3.84
Fourth	6.42 ^a	1.92 ^b	4.17 ^c	4.17	6.33 ^a	1.58 ^b	4.25 ^c	3.96
Sixth	5.75 ^a	2.67 ^b	4.17 ^c	4.21	6.08 ^a	1.83 ^b	3.92 ^c	3.96

Note—Means for a replication within a row with no common superscripts differed at the .05 level by the Newman-Keuls test.

It is worth reiterating that the strength of the data lend strong support to the notion of a "cognitive algebra;" namely, that "Processes for integrating information follow simple rules of ordinary algebra in a wide variety of situations. This cognitive algebra appears to be a general property of the mind, since it is operative in widely different substantive areas" (Anderson, 1974, p. 3). Particularly in regard to the averaging-adding issue, adult subjects might average for a variety of reasons, including application of previously learned cognitive rules (including the concept of an average), or even because they thought an experimenter might expect them to average. However, these sophisticated alternatives are not plausible for a 5-year-old child. The fact that they do indeed average, although having no formal concept of an average, strongly suggests that they are following a cognitive algebra which is an inherent process of the mind. Stated differently and more specifically, averaging appears to be the actual microprocess of information integration, at least for the impression task under standard experimental conditions.

The results for the inconsistent sets are of some interest. One set (good-bad) was used because it virtually represents the archetype of polarity of opposites. Presumably such strong inconsistency should cause severe integration difficulties for small children. However, this was apparently not the case. The children seemed to have little trouble in integrating such disparate information. Previous work has indicated that there is little evidence that adult subjects are bothered by such inconsistency (Hendrick, 1973) and that special experimental arrangements are required before the impression response is affected at all (Hendrick, 1972).

Given such results for both adults and children, one may question whether there is anything that needs to be explained. Perhaps the assumption is wrong—that perceived inconsistency is always unpleasant and must be resolved. The assumption stems from what Rychlak (1968) calls the demonstrative or deductive image of man as a cognitive creature. Rychlak argues persuasively that dialectical reasoning is just as pervasive as demonstrative reasoning. Indeed, the good-bad-neutral data obtained from the children for inconsistent sets is analogous to the thesis-antithesis-synthesis triad of the dialectic. It would be desirable for the future development of cognitive algebras to explore the extent

to which these two (presumed) basic modes of thought are indeed operative in cognitive processing. Developmental research would seem to offer the possibility of providing an answer for this question and others of a related nature.

REFERENCES

- Anderson, N. H. On teaching F instead of t. *Journal of Experimental Education*, 1960, 28, 161-263.
- Anderson, N. H. Averaging versus adding as a stimulus-combination rule in impression formation. *Journal of Experimental Psychology*, 1965, 70, 394-400.
- Anderson, N. H. 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., & Jacobson, A. Effect of stimulus inconsistency and discounting instructions in personality impression formation. *Journal of Personality and Social Psychology*, 1965, 2, 531-539.
- Asch, S. E. Forming impressions of personality. *Journal of Abnormal and Social Psychology*, 1946, 41, 258-290.
- Butzin, C. A., & Anderson, N. H. Functional measurement of children's judgments. *Child Development*, 1973, 44, 529-537.
- Hendrick, C. Averaging vs. summation in impression formation. *Perceptual and Motor Skills*, 1968, 27, 1295-1302.
- Hendrick, C. Effects of salience of stimulus inconsistency on impression formation. *Journal of Personality and Social Psychology*, 1972, 22, 219-222.
- Hendrick, C. Effects of stimulus consistency-inconsistency on impressions of personality. *JSAS Catalog of Selected Documents in Psychology*, 1973, 3, 84-85.
- Hendrick, C., Hoving, K. L., & Franz, C. M. Children's likableness ratings of 22 trait adjectives. *Bulletin of the Psychonomic Society*, 1974, 4, 91-92.
- Peevers, B. H., & Secord, P. F. Developmental changes in attribution of descriptive concepts to persons. *Journal of Personality and Social Psychology*, 1973, 27, 120-128.
- Rychlak, J. F. *A philosophy of science for personality theory*. Boston: Houghton Mifflin, 1968.

NOTES

1. A third concern was with the set size effect, and the original plan for the design included five homogeneous pairs, two HH, one NN, and 2 LL. These data are not reported. There was some difficulty in constructing enough isovalent pairs. Also, for half of the subjects singleton ratings were obtained twice, and for L traits there appeared to be a small regression effect (.28 scale point) from first to second ratings of singletons. Both problems jointly caused some difficulty in interpretation. For these reasons the set size data are not discussed further in this report.

2. Since the traits may be useful in the future, the normative data from the initial ratings have been presented separately (Hendrick, Hoving, & Franz, 1974).

3. It is possible that the choice of the neutral category represented an uncertainty response. However, parsimony suggests that the subjects were just averaging as they did for HM⁺ and LM⁻ pairs. As noted previously, response to inconsistency seems to be minimal under ordinary experimental arrangements (Hendrick, 1972).