

AN EXPERIMENTAL ANALYSIS OF THE CONTRAST EFFECT AND ITS IMPLICATIONS FOR INTERGROUP COMMUNICATION AND THE INDIRECT ASSESSMENT OF ATTITUDE¹

ROBYN M. DAWES,²

University of Oregon, Oregon Research Institute

DAVID SINGER

Oregon Research Institute

FRANK LEMONS

Oregon Research Institute

Experiments I and II were designed to differentiate between alternative interpretations of the contrast effect. Experiment I examines and rejects the hypothesis that the effect is specific to rating scales, since the effect also occurs when subjects are asked to make comparative judgments. Experiment II leads to the rejection of the hypothesis that the effect is purely semantic (i.e., specific to situations in which verbal labels are applied to attitudes); when subjects are asked to write items that would be endorsed by people with whom they agree or disagree, they write much more extreme items for those with whom they disagree than for those with whom they agree. In fact, the items subjects write for people with whom they disagree are not endorsed by those people, the reason most often given being that the items are too extreme. Experiments III, IV, and V demonstrate that the differences are so clear-cut that such attitude statement writing could be used as an indirect measure of attitude. Judges were able to categorize writers' attitudes rather accurately, but accuracy was no greater when the contrast effect was explained to them than when it was not. The implications of these findings for intergroup conflict are discussed, particularly the implication of a positive feedback effect that will lead to exacerbation of group differences.

The *contrast effect* in social judgment (Hovland, Harvey, & Sherif, 1957) refers to an individual's tendency to exaggerate the discrepancy between his own attitudes and the attitudes represented by opinion statements endorsed by people with opposing views, at least when these people are held in low esteem (see Manis, 1966, pp. 71-78, for a review of the literature documenting this effect). To the degree to which this tendency is strong or prevalent, it will exacerbate the conflict between opposing groups; for if members of each group believe that statements of members of the other group represent more extreme attitudes than they in fact do, each group will believe it to be more difficult to

compromise with the other than it in fact would be; further, the belief that members of the other group hold extreme attitudes will lead to a lowering of esteem for that group, which in turn will strengthen the contrast effect. Thus, this effect is both destructive and self-perpetuating.

While the above statements seem intuitively reasonable, it is important to ask exactly what is meant by the phrase "tendency to exaggerate," or, equivalently, to ask how the authors who have investigated this tendency have established its existence. The purpose of the present series of experiments is to examine alternative interpretations that may be placed on previous findings, to collect data meant to discriminate between these interpretations, and finally to develop and demonstrate the validity of an indirect measure of attitude resulting from the earlier studies. Then, the implications of the contrast effect for intergroup communications will be reassessed.

¹ This research was supported in part by National Institute of Mental Health Grant MH-12972 and in part by National Science Foundation Grant GS-1116. The authors are especially indebted to Judy Boylan for her assistance in data collection and analysis.

² Requests for reprints should be sent to Robyn M. Dawes, Oregon Research Institute, P. O. Box 3196, Eugene, Oregon 97403.

In the past, the method used for evaluating the contrast effect has been that of asking subjects to rate the attitude expressed by a particular opinion statement. These ratings have been made with respect to a dimension predetermined by the experimenter (e.g., pro-anti), and they have all involved categorical judgment; that is, the subject is to place the attitude expressed by a particular statement in one of several categories on the specified dimension. This procedure involves two serious problems.

1. The procedure assumes interpersonal comparability of categorical judgment. For example, when a pro individual places a pro statement in a +2 category, whereas an anti individual places it in a +3 category, it is assumed that the categories have the same meaning for the two individuals but that the attitude expressed appears different to them. It could, however, just as easily be assumed that the pro individual and the anti individual assess the attitude equivalently but use the pro and anti categories of rating scales in a different manner. This duality between assuming differential judgment of stimuli and assuming differential use of category boundaries is illustrated in Figure 1. Part A of Figure 1 illustrates the data interpreted according to the assumption that the categories have the same meaning to both pro and anti judges, whereas these judges believe that a statement expressing an opposing attitude represents a more extreme position than does a statement with which they agree. Part B of Figure 1 illustrates the data interpreted according to the assumption that the judges do not believe that the statements represent different extremities of attitude, but each judge allows the extreme categories to cover a broader range of the attitudes with which he disagrees than of those with which he agrees. This latter interpretation is compatible with the current vogue of labeling one's political opposition "extremist." (For a fuller discussion of the duality between interpreting categorical judgments in terms of perception of stimuli and interpreting them in terms of differential use of category boundaries, see Coombs, 1963, or Dawes, 1963.)

2. The procedure requires judgment with

respect to an attribute specified by the experimenter, rather than one chosen by the subject. It is perfectly possible that if the subject were asked to judge which attitude statements are similar, or which he responds to in the same manner (e.g., with approval or anger), the position of the statements on the specified attribute would not be related to such judgments. In other words, the procedure involves *assuming*, rather than demonstrating, that the specified attribute is important in the judge's perception of attitude.

In the context of psychophysics, the philosopher Nelson Goodman (1951) pointed out 20 years ago that procedures which involve judgment with respect to an attribute specified by an experimenter beg the most important perceptual question, which is: What *are* the relevant perceptual dimensions? Hastdorf, Richardson, and Dornbusch (1958) make much the same point in the domain of interpersonal perception when they argue that this phenomenon should be studied more by asking the perceiver to choose his own attributes and categories in describing another person, and less by asking him to rate according to traits selected by the experimenter. The fact that the experimenter himself is a perceiver and a social judge as well as an investigator argues that he will choose important attributes, but even then it is not clear that a judgment concerning the position of stimuli on these attributes will predict other behavior of the judge toward these stimuli. (Of course, the judgment will nevertheless be made—if only because the investigator has asked for such a judgment, and most subjects are polite enough to acquiesce even in tasks that they consider meaningless.)

Furthermore, the procedure of specifying an attribute evaluates the contrast effect only within a semantic framework; that is, it is concerned only with how judges label statements with respect to certain semantic concepts such as pro-anti. It does not follow that judges who label statements differently—even judges using categories in the same way—perceive reality differently, or that one or the other has distorted it.

Experiment I examines the contrast effect by using a comparative, rather than a cate-

one light and one dark patch; in such pairs, one of the members lay on one side of what the subject considered an ideal gray, whereas the other member lay on the other side. (Coombs termed such pairs *bilateral pairs*.)

A judgment of which member of a bilateral pair is more representative of gray can be conceptualized as an extremity judgment. If the dark member appears more extremely dark than the light member appears extremely light, then the subject will say the light member is more representative; if the light member is perceived as more extreme, then the dark one will be judged as more representative; "extremity" is, of course, defined with respect to the subject's conception of ideal grayness.

The present study makes use of such extremity judgments to study the contrast effect.³ Consider two statements, one pro and one anti, that are judged equal in extremity by some hypothetical (and probably non-existent) neutral judge. If an individual who is pro tends to exaggerate the discrepancy between his position and that expressed by the anti statement, he will judge the anti statement to be more extreme; conversely, if an individual is anti, he should judge the pro statement to be more extreme. However, in addition to differences in *categorical* judgments of statements, pros and antis should also *comparatively* rate the statements differently, irrespective of categorization. If the contrast effect were due solely to differences in categorical judgments, then no such differences of extremity judgments should be found when the comparative method is used (see Figure 1b).

It is not possible to obtain statements that are precisely equal in extremity as judged by a neutral individual; further, it was hypothesized that there will be differences in the perception of opinion statements—even by neutral individuals—that exist over and above the hypothesized contrast effect. However, it is possible to test the hypothesized effect on a statistical level. That is, there should be

a *tendency* for pro subjects to judge anti statements as more extreme and anti subjects to judge pro statements as more extreme when making extremity judgments of bilateral opinion statements.

Method

The opinion statements used in this experiment all expressed attitudes toward the war in Vietnam. There were nine pro statements (henceforth referred to as "hawk" statements) and nine anti statements ("dove" statements). These statements, displayed below, can be classified according to three levels of extremity—here labeled "extreme," "moderate," and "mild." The basis for this classification was a series of judgmental studies involving both categorical judgments and unilateral extremity judgments. The final statements, chosen from an initial pool of 48, have the property that virtually no subject *when asked to judge hawk statements alone or dove statements alone* judged a statement in the moderate category to be more extreme than one in the extreme category, or one in the mild category to be more extreme than one in either the moderate or the extreme category.⁴ No bilateral extremity judgments were collected during the item selection procedure.

The subjects used in this experiment were 100 paid University of Oregon students, who were solicited through the campus newspaper during the fall of 1967 and the winter of 1968. The newspaper advertisement specifically asked for hawks and doves on the Vietnam issue, and subjects were accepted in the experiment until there were 50 of each type; the terms hawk and dove were not further defined since their use around campus was quite common, but the newspaper advertisement did specify that some feeling about the Vietnam war was necessary for participation in the experiment. (If any completely neutral subjects lied in order to participate in the experiment, their responses should simply obscure the hypothesized effect.)

The subjects were presented with a questionnaire consisting of 135 pairs of the items displayed above. There are 153 possible pairs, but for reasons that were clearer at the time than they are now, 18 pairs consisting of two hawk statements or two dove statements were omitted. There were two forms of the questionnaire; these forms had the same random order of the 135 pairs, but the order of the two statements within each pair was different on the two forms. The subjects were given the following instructions:

This questionnaire is designed to measure how extreme various statements about the war in Vietnam appear to be. There are 135 parts to this questionnaire, and in each part you will be pre-

³ It should be noted that only the technique has been borrowed from Coomb's (1958) experiment—that the hypothesis tested in this experiment is unrelated to that tested by Coombs.

⁴ A more detailed account of the item selection procedure is available on request from the authors. The procedure is omitted here because of the high face validity of the result.

HAWK STATEMENTS

Extreme

- E1. Only fools and cowards can fail to see that we must defeat the Communists in Vietnam.
 E2. Any and all means are justified in blocking the expansion of Communism in Vietnam.
 E3. No American can in good conscience oppose our efforts in Vietnam; those who do so are, in a word, cowards.

Moderate

- M1. Successful fulfillment of our commitment to Vietnam is necessary if other nations are to respect our national honor.
 M2. As long as the Communists persist in their aggression against South Vietnam, America must resist that aggression.
 M3. Peace can come to Vietnam only if United States military intervention is maintained.

Mild

- m1. Although mistakes are being made by the United States in Vietnam, we seem to be following a reasonable course there.
 m2. The advantages of our Vietnam policy somewhat outweigh its disadvantages.
 m3. Our presence in Vietnam is a necessary evil.

DOVE STATEMENTS

Extreme

- E1. United States intervention in Vietnam is ethically immoral, internationally illegal, and in violation of our own constitution. We should get out now.
 E2. We should get out of Vietnam now, without concern for "face" or prestige, or any other of those nonsensical abstractions with which we justify our barbarity.
 E3. Our involvement in Vietnam is the reflection of a power-mad obsession on the part of a few individuals in our government and is totally without justification.

Moderate

- M1. The United States interference in Vietnam has done little good and much harm.
 M2. There is very little justification for the lives lost in Vietnam.
 M3. Our Vietnam policy is not well thought out, nor is it motivated by any real interest in the people of Vietnam.

Mild

- m1. The American people should become more aware of the cost of our actions in Vietnam.
 m2. The disadvantages of our Vietnam policy somewhat outweigh its advantages.
 m3. The Vietnamese problem should be handled by some international body such as the U. N., not by the United States alone.

sented with two statements. Your job is to tell us which of the statements you judge to be more extreme.

For example, listed below are two statements about the war in Vietnam. Place a check in front of the more extreme statement.

Example

- A. Any and all military means, including nuclear weapons, should be used to stop the threat of Communism in Vietnam.
 —B. The war in Vietnam is a costly but necessary expense for the United States.

Given the above pair, most people would judge Statement A to be more extreme.

In the above example, both statements favored fighting in Vietnam, and hence the choice of which statement was more extreme was reduced to a choice of which statement was more prowar. In the pairs that follow, however, sometimes one statement will be prowar and one statement will be antiwar. Judgments about these pairs may be

more difficult, since what must be judged is whether the pro statement is more pro than the anti statement is anti, or *vice versa*; that is, extremity *per se* must be judged. Please place a check in front of the more extreme statement in every pair. Even if at first glance the two statements seem to be about equally extreme; place a check in front of the one that appears to be the more extreme to you.

Results

The results are presented in Tables 1 and 2; the labels of the rows and columns of these tables refer to the statements displayed above, while the cell entries in the tables indicate the proportion of subjects who judged the column statement to be more extreme than the row statement. Table 1 presents the results for the dove subjects, while Table 2 presents the results for the hawk subjects.

TABLE 1
EXTREMITY JUDGMENTS: PAIRED COMPARISONS—DOVES

Statement	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1. Dm1		71	65	78	84	96	—	86	94	100	—	98	90	100	94	100	100	100
2. Dm3			51	82	69	—	76	71	88	94	86	96	94	100	—	100	100	100
3. Dm2				76	78	88	80	—	96	92	86	98	—	98	98	100	98	100
4. Hm1					58	60	58	60	73	—	78	94	70	94	88	98	—	94
5. Hm2						61	53	53	84	80	73	90	72	—	94	98	98	—
6. DM3							58	47	58	70	78	84	76	78	—	96	98	100
7. DM1								63	66	72	—	82	90	84	98	98	100	100
8. DM2									70	76	76	78	—	82	94	96	96	100
9. Hm3										59	62	—	62	67	78	—	90	98
10. HM1											53	60	56	73	74	92	—	100
11. DE1												53	59	52	88	80	78	94
12. HM3													57	66	71	—	88	100
13. DE2														52	78	78	82	90
14. HM2															60	78	84	—
15. DE3																62	60	80
16. HE3																	65	80
17. HE1																		83
18. HE2																		

Note.—Entries indicate the proportion of people who chose the column statements as more extreme than the row statements. First letter refers to whether statement is hawk or dove, second letter refers to the range (extreme, moderate, etc.), number refers to order of statement (see display of statements in text). For description of boldface entry, see text.

The statements are ranked in these tables according to the principle of stochastic dominance. That is, one statement is ranked above the other if it is judged to be more extreme than the other by 50% or more of the subjects. With a single exception in each table (the boldface entries) this stochastic dominance is transitive; each of these exceptions is

due to one subject. Thus, the ranking of the statements in the two tables is a good representation of how extreme the statements are judged to be by the two groups. This ranking is consistent with the hypothesized contrast effect; the doves have a tendency to judge hawk statements as being more extreme, while the hawks have a tendency to judge dove

TABLE 2
EXTREMITY JUDGMENTS: PAIRED COMPARISONS—HAWKS

Statement	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1. Hm1		50	66	62	70	—	72	72	64	90	86	84	92	—	86	88	92	96
2. Hm2			52	61	71	62	—	80	74	90	88	86	90	90	—	88	94	92
3. Dm1				64	72	66	76	72	78	84	84	—	84	82	82	92	—	90
4. Dm3					66	60	72	72	74	—	84	82	86	84	88	92	86	—
5. Dm2						52	58	64	48	84	—	84	82	74	84	—	90	88
6. HM1							70	58	54	76	72	72	82	—	90	86	88	86
7. HM2								56	62	64	68	92	88	—	80	86	86	86
8. Hm3									—	72	72	72	—	78	78	86	84	88
9. HM3										64	72	66	—	82	92	88	84	92
10. DM3											54	56	66	64	62	70	82	—
11. DM2												58	62	66	72	—	86	86
12. DM1													54	50	68	82	—	94
13. HE3														56	54	64	60	74
14. HE1															54	68	60	82
15. HE2																52	50	55
16. DE2																	62	80
17. DE1																		70
18. DE3																		

Note.—Entries indicate the proportion of people who chose the column statements as more extreme than the row statements. First letter refers to whether statement is hawk or dove, second letter refers to the range (extreme, moderate, etc.), number refers to order of statement (see display of statements in text). For description of boldface entry, see text.

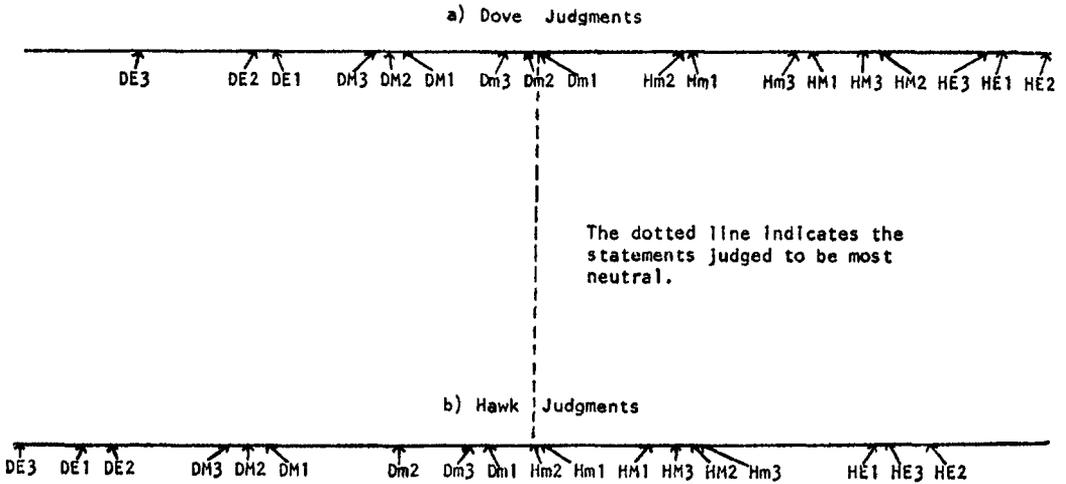


FIG. 2. Comparative judgment representation of dove and hawk judgments.

statements as being more extreme; in fact, with the exception of Hm3 in Table 2, there is a perfect ordering when each category of statements is considered separately; that is, ranking by stochastic dominance, the hawks ranked all of the dove extreme statements as more extreme than any of the hawk extreme statements, while the doves ranked all the hawk extreme statements as more extreme than any of the dove extreme statements, and so on. (It should be noted in passing that the original categorization of the statements is replicated; that is, again with the exception of Hm3, the extreme hawk statements are ranked above the moderate hawk statements, which in turn are ranked above the mild hawk statements; a perfect ordering is found for the dove statements.)

The statements were scaled according to the "double law of comparative judgment" (Coombs, Greenberg, & Zinnes, 1961). This procedure converts the probability with which one statement is judged to be more extreme than another into a distance between the extremity of the two statements, which in turn is converted to a position on the pro-anti continuum; where there were missing comparisons, the incomplete data procedure outlined in Torgerson (1958, pp. 173-176) was employed.⁵

⁵ It was necessary to make a number of assumptions in order to apply the procedure. First, it was assumed that all judgments of which of two dove

The results of this scaling procedure are presented in Figure 2. Figure 2a illustrates the scale positions of the statements based on the doves' judgments, while Figure 2b illustrates these positions based on the hawks' judgments. As is compatible with the earlier analyses, the hawk statements are more extreme than the dove statements on the dove scale, and vice versa.

A statistical test of the contrast effect was devised by noting the number of times hawk statements were judged to be more extreme than dove statements—irrespective of which statements were being considered. The contrast effect predicts that this number should be greater for doves than for hawks.

The mean number of hawk statements chosen as more extreme by the hawks was 32.94, while the mean number of hawk statements chosen as more extreme by the doves was 58.64. A *t* test evaluating the difference between these two means yielded a value of -10.01, which is significant well beyond the .001 level (*df* = 98).

statements or which of two hawk statements was more extreme were unilateral judgments, whereas the judgments comparing a hawk with a dove statement were bilateral. Second, the distribution of the concept "neutrality" was assumed to have a variance equal to that of the distributions of the statements, which, in turn, were all assumed to have equal variance and to be uncorrelated. Finally, the absolute zero point (neutrality) was identified with the statement scaled at least extreme.

Discussion

Extremity, of course, makes sense only with respect to what is considered neutral. One possible interpretation of the results is that hawks and doves have a different idea of what represents a neutral position toward the war in Vietnam. Upon close inspection, this interpretation is seen not to be an alternative to the interpretation that hawks and doves have different ideas about extremity. That is, to say that two people judge neutrality differently is equivalent to saying that they judge extremity differently; extremity means, by definition, distance from neutrality, and what the present data indicate is that a given statement's distance is judged differently by doves and hawks, in a direction predicted by the contrast effect.

The interpretation of the finding in terms of neutrality is compatible with adaptation-level theory (Helson, 1964). Granted such factors as selective exposure and attention, it is reasonable to suppose that a pro has more contact with pro attitudes and an anti with anti attitudes—which would result in the pro's subjective neutral point to be more pro and the anti's more anti. The results reported in Tables 1 and 2 and Figure 2 would thus follow.

Even if the contrast effect cannot be explained purely in terms of differences in categorical judgment, however, it may be specific to judgments of statements' positions on a pro-anti continuum. If it were, it would be of restricted interest—except to investigators concerned with semantic judgment and labeling. If, on the other hand, it can be shown to represent a genuine distortion of the position of people holding opposing attitudes, it is of more general social-psychological interest, and it has the implication for intergroup communication and conflict discussed at the beginning of this paper.

EXPERIMENT II

To evaluate whether the contrast effect is specific to verbal labeling, the usual procedure for studying it was reversed. That is, rather than having subjects categorize or compare opinion statements constructed by the experimenter, the subjects were asked to generate

the opinion statements themselves. Specifically, subjects were asked to construct opinion statements that they thought would be endorsed by the "typical" hawk or dove on campus. If the contrast effect were purely a labeling phenomenon, the statements so generated would be neither systematically more extreme nor less extreme as a function of the attitude of the writer. If, however, the contrast effect is due to genuine social distortion, subjects writing statements to be endorsed by people holding opposing views should write more extreme statements than those they write when they are writing statements to be endorsed by people with whom they agree. Or, finally, there is the possibility that they write less extreme statements; for, if a subject is to be consistent with his judgment that statements with which he disagrees express extreme attitudes, he should write very mild statements if he wishes to write statements that are—for him—moderate. These three possibilities can be stated as conflicting hypotheses.

1. *The semantic hypothesis.* The contrast effect is specific to categorizing or comparing statements on a pro-anti continuum. If this hypothesis is correct, statements written to be endorsed by the typical pro or anti individual should not vary systematically as a function of the writer's own attitude.

2. *The person displacement hypothesis.* The contrast effect occurs because individuals exaggerate the differences between their own attitudes and the attitudes of those with whom they disagree. It follows that an individual attempting to write a statement to be endorsed by the typical person holding an opposing attitude will write a more extreme statement than he will when he attempts to write a statement to be endorsed by the typical person with whom he agrees.

3. *The statement displacement hypothesis.* As indicated by Experiment I, people who disagree with a statement tend to regard it as more extreme than do people who agree with it. It follows that *if* people do not disagree about the extremity of their own and others' positions, but only exaggerate the extremity of *statements*, then a person writing a statement that is meant to characterize the opinion

of someone with whom he disagrees should write a statement that would sound rather moderate to others (e.g., since hawk *statements* seem more extreme to doves than to hawks, the dove wishing to produce what he regards as a moderate statement should generate one that is less extreme than one that would be generated by a hawk; if a person sees a statement as extreme, he is less likely to attribute it to another person than is someone who sees it as moderate).

One way to evaluate "the extremity" of the statements an individual writes is to ask people whether they will endorse these statements. If either the person displacement hypothesis or the statement displacement hypothesis is correct, the individual will write statements that are not in fact endorsed by people holding opposing views. If the person displacement hypothesis is correct, the statements not endorsed will be rejected because they are too extreme, whereas if the statement displacement hypothesis is correct, they will be rejected because they are too mild.

Method

Again the issue studied was the war in Vietnam. Twenty hawks and 20 doves, responding to a newspaper advertisement similar to that described in Experiment I, were asked to write four statements that they thought would be endorsed by the typical hawk on campus and four statements that they thought would be endorsed by the typical dove on campus. The actual instructions read as follows:

We are planning to engage in research concerning attitudes toward the Vietnamese war. We are going to use college students in this research, and we will be asking them to endorse or reject the sorts of statements that college students typically make about the war. We are asking your help in obtaining such statements. Please write four statements that you think would be endorsed by the typical hawk on campus. . . . Now, please write four statements that you think would be endorsed by the typical dove on campus.

Questionnaires were constructed from the statements that the subjects wrote. Two of the statements that the subject thought would be endorsed by the typical hawk and 2 he thought would be endorsed by the typical dove were randomly selected from each subject's statements. The resulting pool consisted of 40 statements written by doves (henceforth referred to as dove-written statements) for endorsement by the typical dove, 40 statements written by hawks (hawk-written statements) for endorsement by the typical hawk, and

40 hawk-written statements for endorsement by the typical hawk. Two questionnaires were formed from these statements, the first consisting of the 80 statements written for endorsement by the typical hawk, and the second consisting of the 80 statements written for endorsement by the typical dove; these statements were ordered randomly on the questionnaires.

Twenty-five hawks and 25 doves were recruited in the usual manner.⁶ The hawks were presented with the questionnaire made up of statements to be endorsed by the typical hawk, and the doves were presented with the questionnaire constructed of statements to be endorsed by the typical dove. The subjects were asked to endorse or reject each statement, and, if they rejected the statement, to state whether their reason for doing so was that the statement was too strongly worded or too mildly worded. The actual instructions read as follows:

Listed below are 80 statements which various people have made about the Vietnam war. Your task is to place a check (✓) in front of the statements with which you agree. If you do *not agree* with the statement because you feel it is too strongly worded, place an S in front of the statement. If you do *not agree* with a statement because you feel that it is too mildly worded, place an M in front of that statement. If you do not agree with the statement for any other reason; i.e., you feel that it is too ambiguous, you don't like the way it is worded, etc., place an X in front of the statement; then please go back and try to determine whether the *general attitude* expressed by the statement is too strong or too mild; if it is, place an S or an M, respectively, in parentheses after your X mark; if not, place a ✓ in parentheses.

Results

The number of hawk-written and dove-written statements rejected by each subject and the reason for rejection were tabulated. The semantic hypothesis predicts that there would be no difference in the number of hawk-written statements and dove-written statements rejected by hawks and doves. Both the person displacement hypothesis and the statement displacement hypothesis predict that the hawks should reject more dove-written statements and the doves more hawk-written statements, and, *equally important*, they predict opposite reasons for rejection. According to the person displacement hypothesis, the predominant reason should be that the statements are too extreme; according to the statement

⁶ Great care was taken so that no subject participated in more than one phase of this experiment.

TABLE 3

DOVE AND HAWK REJECTION OF HAWK-WRITTEN AND DOVE-WRITTEN STATEMENTS AND REASON FOR REJECTION

Statement	Reason	
	Too strong	Too mild
Dove rejection		
Dove written	7.84	3.00
Hawk written	9.32	3.36
Hawk rejection		
Dove written	15.68	1.44
Hawk written	11.40	2.00

displacement hypothesis, the predominant reason should be that they are too mild.

Table 3 presents the mean number of dove-written and hawk-written statements rejected by doves and hawks together with the reason for rejection. The pattern of the means supports the person displacement hypothesis. That is, doves reject more hawk-written statements, and the hawks reject more dove-written statements; further, the reason for the vast majority of rejections is that the statements are too strong.

In order to simplify the statistical analysis, the raw data of this study were converted to difference scores. That is, a 2 x 2 analysis of variance was conducted based on the following independent variables: endorser's attitude (hawk versus dove) and reason for rejection (too strong versus too mild). Within each of the resulting four cells (see Table 4), the observations for each respondent (endorser) consisted of the number of hawk-written statements he rejected minus the number of dove-written statements he rejected. In this analysis, the hypothesis predicting that hawks should reject more dove-written statements and doves more hawk-written statements would emerge as a main effect, with the

TABLE 4

NUMBER OF HAWK-WRITTEN STATEMENTS REJECTED MINUS NUMBER OF DOVE-WRITTEN STATEMENTS REJECTED PATTERN OF MEANS TABLE

Group	Reason	
	Too strong	Too mild
Doves	1.48	.36
Hawks	-4.28	.56

TABLE 5

ANALYSIS OF VARIANCE TABLE

Source of variation	SS	df	MS	F
Between subjects	600.41	49		
Groups (A)	193.21	1	193.21	22.78*
Error	407.20	48	8.48	
Within subjects	708.50	50		
Reasons (B)	86.49	1	86.49	10.38*
A x B	222.01	1	222.01	26.65*
Error	400.00	48	8.33	

* $p < .01$.

prediction being that the difference scores for hawks would be lower than the difference scores for doves. The person displacement hypothesis predicts an interaction effect as well; the difference scores will be higher for doves and lower for hawks when the reason for rejection is that the statements are too extreme and vice versa, whereas the statement displacement hypothesis makes the reverse prediction.

The means of the difference scores are presented in Table 4; as follows from Table 3, the direction of these means supports the person displacement hypothesis. First, the overall means are higher for doves than for hawks (i.e., relative to the number of dove-written statements rejected, doves rejected more hawk-written statements than did hawks), and the effect is much stronger when the reason given for rejection was that the statement was too strong than when the reason given was that the statement was too mild.

An analysis of variance is summarized in Table 5. The predicted difference between the groups and the interaction between groups and their reasons for rejecting statements are both highly significant.

Note that the reasons factor is also significant; across both groups there was a tendency to reject dove-written statements when the reason was that the statements were too extreme and hawk-written statements when the reason was that the statements were too mild.⁷

⁷ A speculation: since the time of this study, the prevailing mood in the country seems to have switched from one of hawkishness to one of dovishness; perhaps this interaction reflects a greater commitment or emotional involvement on the part of the doves than on the part of the hawks.

Discussion

The results of this experiment are inconsistent with the semantic hypothesis and the statement displacement hypothesis; they are consistent with the person displacement hypothesis. This person displacement interpretation is particularly striking when viewed in conjunction with the results of Experiment I, for not only do subjects write statements that are too extreme to be endorsed by people with opposing attitudes (Experiment II), but, in addition, subjects tend to judge such opposing statements to be extreme (Experiment I). (The hawk who refuses to endorse a dove-written statement because he judges it to be too extreme is refusing to endorse a statement that may be seen as even more extreme by the dove who wrote it than by himself.)

Further checks could, of course, be run; for example, it would be possible to present judges with hawk-written statements meant to be endorsed by the typical hawk and dove-written statements meant to be endorsed by the typical hawk and ask them to judge which items are more extreme. Such a procedure was deemed unnecessary. The differences are so clear-cut that it was decided to explore the possibility of classifying people on the basis of their statement writing.

EXPERIMENT III

Selltiz, Edrich, and Cook (1965) have suggested that differences in categorical ratings can lead to an indirect measure of the rater's own attitude. Since, in general, an item is given a more extreme rating by a person who disagrees with it than by a person who agrees with it, the more extreme the rating given to an item, the higher the probability that the rater disagrees with it. These authors collected evidence that supported their suggestion, but the effect appears to have been too slight for the raters in their study to be categorized with much accuracy.⁸

⁸ This is our judgment derived from examining the data presented by the authors; they themselves did not treat rating as the independent variable and the rater's attitude as the dependent variable, but adopted the more traditional approach of their predecessors of treating the rater's attitude as the independent variable and the rating as the dependent

In Experiment II, which supported a perceptual interpretation of the contrast effect, subjects generated more extreme statements when writing for endorsement by the typical person with whom they disagreed than when writing for endorsement by the typical person with whom they agreed. The purpose of this experiment was to determine how well statement writing could be used to classify attitudes of writers, more specifically, how well psychologists could classify the hawks and doves of Experiment II on the basis of the attitude statements they wrote.

Method

Six PhD psychologists from the Oregon Research Institute (one of whom was a visitor from the Netherlands) and three graduate research assistants in psychology served as judges. They were presented with the eight statements generated by each of the writers; the four statements written for endorsement by the typical hawk were typed on the left side of a large index card, and those for endorsement by the typical dove were typed on the right side. Each judge was asked to separate the 40 cards into two piles of 20 cards each—one pile for the statements written by the hawk subjects and one for the statements written by the dove subjects. The basic principle for sorting was explained as follows.

If, in fact, people have a tendency to exaggerate the differences between themselves and those with whom they disagree, then when they are asked to write items typical of those with whom they disagree, they should write relatively extreme items; when they are asked to write items typical of those with whom they agree, they should write less extreme items. Hence, these items could be used to assess their attitudes.

Results

All nine judges were able to classify the statement writers as hawks or doves with over 50% accuracy ($p < .01$, by sign test). The average percentage of correct classification was 65%.

The consensus judgments were 78% accurate, as indicated in Table 6, which presents the contingency between consensus judgment and actuality ($\chi^2 = 10.02$, $p < .01$).⁹

one; thus, while their study constitutes a replication of earlier findings, it is not certain how well the ratings they obtained could have been used to classify the attitude of the rater.

⁹ All statistical tests reported include a "correction for continuity."

TABLE 6

Consensus judgment	Actual classification	
	Hawk	Dove
Hawk	15	4
Dove	5	16

Discussion

The judges were accurate. Analyzing the results across statement writers and within judges, we find that all of the judges are more accurate than chance expectation. Analyzing the results within statement writers and across judges, we find that a significant proportion of the statement writers are categorized accurately by consensus judgment.

EXPERIMENT IV

The purpose of Experiment IV was to replicate the results of Experiment III using the statements written by Wallace supporters and the supporters of the write-in protest. The procedure was almost identical to that of Experiment III *mutatis mutandis*—with the exception that two different sets of statements were used. Just prior to the 1968 elections, 30 supporters of the write-in protest and 15 Wallace supporters were recruited for statement writing in a manner similar to the hawk-dove procedure. Statement Set 1 consisted of the statements written by the 15 Wallace supporters and statements written by a randomly selected group of 15 of the supporters of the write-in protest. Set 2 consisted of the statements written by the same 15 Wallace supporters and those written by the remaining 15 supporters of the write-in protest. A tenth judge, a terminal graduate student in psychology, was used so that five judges sorted each set of statements.

Results

All 10 individual judges were able to classify the writers as Wallace supporters or supporters of the write-in protest with over 50% accuracy ($p < .01$ by sign test). The average percent accuracy was 77%.

The consensus judgments were 87% accurate, as indicated in Table 7, which presents the contingency between the consensus judge-

TABLE 7

Consensus judgment	Actual classification	
	Wallace supporters	Supporters of the write-in protest
Wallace supporters	12	3
Supporters of the write-in protest	3	27

ment and the actual category to which the writer belonged ($\chi^2 = 19.01$, $p < .001$).

Discussion

The improvement in accuracy in Experiment IV as compared to Experiment III is probably due to the fact that the attitudinal differences between Wallace supporters and supporters of the write-in protest were greater than those between hawks and doves.

EXPERIMENT V

In Experiments III and IV, the judges were told about the contrast effect. The purpose of this experiment was to determine whether naive judges could categorize writers with above-chance accuracy, and whether accuracy of categorization is greater when judges are informed of the contrast effect than when they are not. The statements written by the Wallace supporters and the supporters of the write-in protest were used. Judges were divided into two groups, one of which was informed about the contrast effect and one of which was not. It seemed reasonable to hypothesize that judges would do better when told of the effect.

Method

Judges in this study were drawn from two different populations. Sixty-two were students at the University of Oregon during the summer of 1970 and were paid for their participation. The other group of judges were drawn from a slightly more sophisticated population; it consisted of 19 employees of the Oregon Research Institute—including secretaries, research assistants in fields other than psychology, computer programmers and operators, and various other employees with the exception of professional psychologists.

The procedure used for *both* groups follows, although the data are treated separately. The judges were divided randomly into two groups; an *informed* group received roughly the same instructions as in

TABLE 8

Consensus judgment	Actual classification	
	Wallace supporters	Supporters of the write-in protest
Wallace supporters	12	6
Supporters of the write-in protest	3	24

Experiment IV, while an *uninformed* group had the explanation of the contrast effect omitted. In the student sample, a randomly selected half of each group was asked to sort the writers in Set 1, the remainder in Set 2. In the Oregon Research Institute sample, only Set 1 was used.

Results

All but 5 of the 62 student judges were able to classify the writers as Wallace supporters or supporters of the write-in protest with over 50% accuracy ($p < .001$, by sign test). The average accuracy was 64%.

The consensus judgments were 80% accurate, as indicated in Table 8, which presents the contingency between consensus judgment and actual classification ($\chi^2 = 12.60$, $p < .001$).¹⁰

The performance of the judges in the informed group was virtually identical to that of the judges in the uninformed group. Individual judges in both groups averaged 64% accuracy; the consensus accuracy of the informed group was 78%, while that of the uninformed group was 76%.

Eighteen of the 19 Oregon Research Institute judges were able to classify the writers as Wallace supporters or supporters of the write-in protest with over 50% accuracy ($p < .001$, by sign test). The average accuracy was 67%.

The consensus judgments were 87% accurate, as indicated in Table 9, which presents a contingency between consensus judgment and actual classification ($\chi^2 = 13.33$, $p < .001$).

Again, however, the performance of the two groups was virtually identical. Individual

¹⁰ Since 62 student judges classified each of the Wallace supporters, it was necessary to run an additional judge in order to obtain the odd number of judges necessary to guarantee a consensus judgment.

TABLE 9

Consensus judgment	Actual classification	
	Wallace supporters	Supporters of the write-in protest
Wallace supporters	13	2
Supporters of the write-in protest	2	13

judges in the informed group averaged 67% correct, as did those in the uninformed group. Further, the consensus judgment of the informed group was 78%, while that of the uninformed group was 70%.¹¹

Discussion

The results do not support the hypothesis that awareness of the contrast effect increases the accuracy of judgment. Many subjects reported difficulty in performing the task, and accuracy was somewhat lower than that of psychologists in Experiment IV. On the average, the psychologists were 77% correct, whereas the other 83 judges on the average were only 65% correct. The number of writers correctly classified by the psychologists and other judges differed significantly ($t = 2.65$, $p < .01$). (A t test in this context actually lacks power, since it fails to control for the fact that the same writers were being classified by the two sets of judges; a more powerful test, however, could only increase the significance level.)

The judges in the uninformed group in this experiment were informed of the contrast effect after they had completed their sorting. Many of these judges reported that they had used their own version of the contrast effect in making these judgments. Thus, a paradox arises: If people naively understand that people tend to exaggerate the differences between themselves and those with whom they disagree, why do they continue to do it?

At any rate, the success of the psychologists in classifying the 40 hawk and dove statement writers and the 45 writers who

¹¹ Since the judges themselves were equally accurate in each group, there is little reason to believe that this consensus difference would be statistically reliable were a larger sample collected.

supported Wallace or the write-in protest is quite impressive. The success of the other two groups in classifying the latter group of 45 subjects is somewhat less impressive, but is nevertheless well above chance expectation. The success rate of all our judges compares very favorably, for example, with the success of expert clinical psychologists in attempting to differentiate psychotic from neurotic patients on the basis of their MMPI profiles (Goldberg, 1965); these judges also were attempting to differentiate two groups in a population with an approximately 50-50 base-rate split.

The subject whose attitude is to be assessed is asked to do something very straightforward: to write, to the best of his ability, statements that would be endorsed by the typical person with whom he agrees and the typical person with whom he disagrees. The statements he writes are not only indicative of his own attitude, but in contexts such as those of the present studies, may be used to categorize him with a high degree of accuracy. Statement writing might, then, be an effective means of indirectly assessing attitude (if future studies show it works in other contexts), as well as one that does not involve any serious deception, or infringement on the privacy of the person whose attitude is being assessed.

GENERAL DISCUSSION

The conclusion from Experiments I and II is that the contrast effect is not a labeling phenomenon; subjects in the present experiments had a strong tendency to exaggerate the differences in attitude between themselves and the people with whom they disagree. Experiments III, IV, and V demonstrate that the differences are so clear-cut that it is even possible to classify people on the basis of their statement writing with a high degree of accuracy.

There are many possible explanations for the contrast effect. Perhaps we have more difficulty than we are aware of in assimilating social information and hence "gate out" (Posner, 1964) incoming information we believe to be irrelevant—that is, information indicating moderation or neutrality (see Manis, Gleason, & Dawes, 1966). This dif-

ficulty-based explanation is consistent with the finding that while we may process huge amounts of information when unconsciously performing habitual tasks, our capacity for consciously processing information in new tasks is severely limited (see Fitts & Posner, 1967), and forming impressions about attitudes is generally a conscious task; the difficulty may further be exacerbated by the fact that we form many impressions about attitude on the basis of what we read, and we may be much worse at understanding what we read than we think we are (see Bartlett, 1932; Dawes, 1964, 1966). In contrast to the information-processing explanation, Manis (1961) has proposed a motivational explanation, which holds that pressure to change our own attitudes is reduced by displacing the attitude of the person with whom we disagree.

As noted at the beginning of this paper, the contrast effect may be self-perpetuating—*functionally autonomous* in Allport's (1937) terminology—because it may well lead to a lowering of esteem for those holding opposing views, and such lowered esteem may enhance the effect. Moreover, it is detrimental to cooperation and compromise.

A closely allied source of intergroup conflict that may also be self-perpetuating is one termed *autistic hostility* by Newcomb (1947). Briefly, autistic hostility is the process of avoiding contact with opposing groups and hence avoiding information that may serve to reduce intergroup conflict. To quote Newcomb, Turner, and Converse (1965):

If the members of one group share hostile attitudes toward another group, they are quite likely to take steps to reduce communication with that group. Such behavior tends to maintain the hostility, because of reduced opportunities for the kind of unlearning that is necessary if the group norms are to prescribe less rather than more hostility [p. 444].

Moreover, the contrast effect and autistic hostility may even enhance each other. The distortion involved in the contrast effect may increase hostility, which may in turn reduce communication, which may in turn exacerbate selective exposure and attention to extreme statements, which may in turn enhance the contrast effect.

The present research indicates that the contrast effect involves a distortion of reality,

rather than a particular use of verbal labels. Further, it may be autonomous, and perhaps even result in a self-amplifying conflict between groups. As has been pointed out by Maruyama (1963), such positive feedback phenomena (which he terms "deviation amplifying") may be found in a wide variety of social and clinical situations. Further, subsystems that may in themselves involve negative feedback (the contrast effect may lead to equilibrium within the individual) can be embedded in larger systems that involve positive feedback (the resulting group conflict). The research reported here suggests that the contrast effect may be part of such a system.¹²

¹² Since the time that this article was written, the basic effect has been replicated in two additional contexts. Results are reported in Dawes, MacPhillamy, and Singer (1971).

REFERENCES

- AGER, J. W., & DAWES, R. M. The effect of judges' attitude on judgment. *Journal of Personality and Social Psychology*, 1965, 1, 533-538.
- ALLPORT, G. W. *Personality*. New York: Holt, 1937.
- BARTLETT, F. C. *Remembering: A study in experimental and social psychology*. Cambridge, England: Cambridge University Press, 1932.
- COOMBS, C. H. On the use of inconsistency of preferences in psychological measurement. *Journal of Experimental Psychology*, 1958, 55, 1-7.
- COOMBS, C. H. Some symmetries and dualities among measurement data matrices. In *Thurstone Hall dedication conference*. New York: Holt, Rinehart & Winston, 1963.
- COOMBS, C. H., GREENBERG, M., & ZINNES, J. L. A double law of comparative judgment for the analysis of preferential choice and similarities data. *Psychometrika*, 1961, 26, 165-171.
- DAWES, R. M. The logic of S-R matrices. *Psychological Review*, 1963, 70, 365-368.
- DAWES, R. M. Cognitive distortion. *Psychological Reports*, 1964, 14, 443-459.
- DAWES, R. M. Memory and distortion of meaningful written material. *British Journal of Psychology*, 1966, 57, 77-86.
- DAWES, R. M., MACPHILLAMY, D. J., & SINGER, D. Studies on emotional involvement and the distortion of social perception. Unpublished research report, *Oregon Research Institute Research Bulletin*, 1971.
- FITTS, P. M., & POSNER, M. I. *Human performance*. Belmont, Calif.: Wadsworth, 1967.
- GOLDBERG, L. R. Diagnosticians vs. diagnostic signs: The diagnosis of psychosis vs. neurosis from the MMPI. *Psychological Monographs*, 1965, 79(9, Whole No. 602).
- GOODMAN, N. *The structure of appearance*. Cambridge: Harvard University Press, 1951.
- HASTDORF, A. H., RICHARDSON, S. A., & DORNBUSCH, S. M. The problem of relevance in the study of person perception. In R. Tagiuri & L. Petrullo (Eds.), *Person perception and interpersonal behavior*. Stanford: Stanford University Press, 1958.
- HELSON, H. *Adaptation-level theory: An experimental and systematic approach to behavior*. New York: Harper & Row, 1964.
- HOVLAND, C. I., HARVEY, O. J., & SHERIF, M. Assimilation and contrast effects in reaction to communication and attitude change. *Journal of Abnormal and Social Psychology*, 1957, 55, 244-252.
- KELLY, H. H., HOVLAND, C. I., SCHWARTZ, M., & ABELSON, R. P. The influence of judges' attitudes on three methods of attitude scaling. *Journal of Social Psychology*, 1955, 42, 147-158.
- MANIS, M. The interpretation of opinion statements as a function of recipient attitude and source prestige. *Journal of Abnormal and Social Psychology*, 1961, 63, 82-86.
- MANIS, M. *Cognitive processes*. Belmont, Calif.: Wadsworth, 1966.
- MANIS, M., GLEASON, T. C., & DAWES, R. M. The evaluation of complex social stimuli. *Journal of Personality and Social Psychology*, 1966, 3, 404-419.
- MARUYAMA, M. The second cybernetics: Deviation amplifying mutual causal processes. *American Scientist*, 1963, 51, 164-179.
- NEWCOMB, T. M. Autistic hostility and social reality. *Human Relations*, 1947, 1, 69-86.
- NEWCOMB, T. M., TURNER, R. H., & CONVERSE, P. E. *Social psychology: The study of human interaction*. New York: Holt, Rinehart & Winston, 1965.
- POSNER, M. I. Information reduction in the analysis of sequential tasks. *Psychological Review*, 1964, 71, 491-504.
- SELLTIZ, C., EDRICH, H., & COOK, S. W. Ratings of favorableness of statements about a social group as an indicator of attitude toward the group. *Journal of Personality and Social Psychology*, 1965, 2, 408-415.
- TORGERSON, W. S. *Theory and methods of scaling*. London: Wiley, 1958.

(Received July 27, 1970)