A COMPARATIVE STUDY OF INDIVIDUAL, MAJORITY, AND GROUP JUDGMENT

DEAN C. BARNLUND
Northwestern University

THE comparative quality of decisions made by groups and by individuals working alone has been tested under a wide variety of experimental conditions. In general, these studies indicate that group judgments are superior to individual judgments on certain types of intellectual problems (2, 6). Where experiments have employed groups composed of persons of different levels of ability, however, it is not clear whether the quality of the decisions is due to the greater influence of the more capable members of the group or is a specific consequence of group thinking itself. Do groups make better decisions because the less intelligent capitulate to the more intelligent members? Or are there psychological factors inherent in group interaction which produce the higher level of performance? When each group member possesses unique information or ideas it is not unreasonable to expect that interaction will increase the total amount of information and enlarge the perspective of the group as a whole. But what happens to the level of group judgments when interaction occurs among persons who are equally informed and talented?

The present investigation is concerned with how decisions made by individuals working alone compare with the pooling of individual judgments through majority vote and with decisions reached through the process of group discussion when: (*a*) the membership of the group is homogeneous with respect to ability to solve the assigned problem; (*b*) the task is complex, couched in prejudicial terms, and involves a range of possible solutions; and (*c*) individuals and groups are permitted the same length of time to complete their tasks. Finally, the study seeks to determine some of the factors that account for any differences observed in individual and group performance.

METHOD

Subjects

The *S*s used in this experiment were students enrolled in freshman courses in group discussion over a three-year period at Northwestern University. The members of eight classes were used, 174 students in all. Of these, 143 were assigned to experimental groups, and the remaining 31 served as control *S*s.

Procedure

At the first meeting of the classes, Form A or Form B of the "Recognition of Valid Conclusions" test was administered. Form A and Form B were alternated as the first and final measures of problem-solving ability throughout the experiment to reduce any biasing effects growing out of differences in the two forms. Each student was given a copy of the test and an answer sheet and instructed to work out his solutions to the 30 problems individually. *S*s were given the 50-minute period to complete the test items.

Each member of the class was then ranked according to the total number of items he answered correctly on the first form of the test. Eight or nine weeks later, before the end of the academic quarter, experimental groups were created. Four or five groups were formed in each of the classes used in the experiment. All students who received the same or similar scores on the first test were placed together so that homogeneous groups were created. Experimental groups were then given a single answer sheet and copies of the alternate form of the test and instructed to reach a group decision on each of the 30 problems. Experimental and control *S*s were again given 50 minutes to finish the test. Members of the control classes repeated the test under the original conditions, solving items on the alternate forms of the test individually. A total of 29 experimental groups participated in the experiment.

The final 10 group sessions were tape recorded in their entirety. An analysis was made of each of the 30 decisions reached by the 10 groups to isolate the specific kinds of mistakes that contributed to the majority of group errors. Following all group tests, discussions were held with the experimental *S*s concerning the factors they felt influenced their performance as members of the groups.

Problem

Many investigators of group phenomena have admitted difficulty in finding or constructing suitable instruments for testing the efficiency and accuracy of group decision-making. Problems, to be realistic, should be complicated enough so they cannot be solved by intuition. They should be sufficiently difficult to test the limits of individual and group thinking. Social problems normally can be solved in a variety of ways, and test problems should contain this same feature. The difference between a right and wrong decision, however, should be clear and demonstrable. If possible, problems should be presented so as to involve the total

personality of the individual and permit his prejudices to influence his judgment as they do in a majority of everyday problem-solving experiences.

The instrument used in this experiment, the Bradley test of *Formal Validity in Problem Solving*, seemed particularly well adapted for this purpose (1). The first section of the test entitled, "Recognition of Valid Conclusions," proved long enough and sensitive enough to provide data on the experimental hypothesis. The 30 problems which make up the test consist of partially constructed arguments of varying degrees of difficulty. Two statements are given which are to be assumed to be materially true. The problem is to select the conclusion that follows most logically from the premises. The arguments cover a wide range of subjects and are phrased deliberately to complicate the decision for the reader; that is, statements involve atheists, Communists, Republicans, college professors, and other terms likely to prejudice judgment. An example of one of the problems is given below:

Some Communists are advocates of heavy taxes;
All advocates of heavy taxes are conservative Republicans;
Therefore:

a. Some advocates of heavy taxes are not Communists
b. Some Communists are conservative Republicans
c. Some conservative Republicans are Communists
d. Some Communists are advocates of heavy taxes
e. None of these conclusions follows

The validity and reliability of the instrument has been established. The 30 problems on each of the two forms include the 19 valid moods of the syllogism along with the 11 most common fallacies. The test has been successful in discriminating among college students with different backgrounds in logic, mathematics, and problem solving. Intercorrelation of the two forms yields a raw score "*r*" of .85 (PE, .015) and a weighted "*r*" of .88 (PE, .012). Items have been carefully scaled in the final forms so that similar scores represent similar patterns of individual errors.

RESULTS

Measures of the relative effectiveness of individual, majority, and group judgments were obtained from scores made on the two forms of the "Recognition of Valid Conclusions" test.

The number of items answered correctly on the first form was used to set up homogeneous groups and to determine the level of ability represented by the average scores and "superior" scores of members of the experimental groups when working alone.[1] The relative accuracy of problem solving under conditions of majority rule was derived from an item analysis of the individual answers of each group member. This "mathematical majority" indicated how the groups would have scored if they had pooled their opinions by secret ballot. Of the total of 829 decisions made by the experimental groups, 22 were found to be deadlocks. These occurred whenever a group of four or six *S*s divided their votes equally between right and wrong answers. The results of splitting these decisions evenly and from crediting all of them to the advantage of the majority are recorded in Table 1 under "Deadlocks divided" and "Deadlocks credited." The quality of group thinking was measured by computing the mean scores of experimental groups on the second form of the test when they were required to reach consensus on each of the test problems.

The mean scores obtained under the various experimental conditions and the *t* values they yield are summarized in Table 1. The average scores of members of the experimental groups working alone are not significantly larger or smaller than the mean of majority scores when the 22 deadlocks are counted as correct in half of the instances and incorrect in the other half. When all deadlocks are resolved in favor of the correct decision, majority rule proves to be superior to the average performance of the individual group members. The "superior" members of the experimental groups, on the other hand, did significantly better than the majority when deadlocks were split, and as well as the majority when deadlocks were counted as correct solutions to the problems.

Group decisions were found to be clearly superior to individual decisions. As a result of discussion, experimental groups obtained mean scores that were significantly higher, at the .01 level, than "superior" members of the same groups were able to attain through individual effort. These findings also hold true when results for Form A and Form B are analyzed separately. Groups whose members scored initially near the upper limit of the test, 28 or 29 correct answers out of a possible 30, gained least from solving problems cooperatively. The largest gains were made by groups whose initial scores were low although nearly all of the experimental groups, with the exception of the highest scoring group in each class,

[1] The "superior" member of a homogeneous group is something of a misnomer. Experimental groups were made up of *S*s whose initial scores differed by no more than a few points. In each case the "superior" member refers simply to the individual who made the highest individual score in the group despite its homogeneous character.

TABLE 1

COMPARISON OF INDIVIDUAL, MAJORITY, AND GROUP SCORES ON THE "RECOGNITION OF VALID CONCLUSIONS" TEST

				t
Individual Decisions		Majority Decisions		
Means of average individual scores	17.5	"Deadlocks divided"	17.9	1.73
		"Deadlocks credited"	18.3	2.72*
Means of "superior" individual scores	18.8	"Deadlocks divided"	17.9	3.23**
		"Deadlocks credited"	18.3	1.66
Individual Decisions		Group Decisions		
Means of average individual scores	17.5	.ean scores of groups	21.9	9.46**
Means of "superior" individual scores	18.8			5.77**
Majority Decisions		Group Decisions		
"Deadlocks divided"	17.9	Mean scores of groups	21.9	6.60**
"Deadlocks credited"	18.3			5.95**

* Significant at .05 level.
** Significant at .01 level.

made substantial gains as a result of group deliberation. Students in the lowest fifth of their classes as a group often rivalled the performance of the most brilliant member of the class working alone. In only two of the 29 experimental groups did students working together fail to outperform their own best member.[2]

When majority rule is compared with group consensus, the results show a similar large and significant advantage for group decision-making. Crediting all deadlocks from divided votes reduces the size of the group advantage over majority decisions, but its value is still highly significant.

The 31 control *S*s had mean scores on the initial administration of the "Recognition of Valid Conclusions" test of 18.5. (Control *S*s made an initial mean score of 18.8 on Form A and of 18.2 on Form B.) On the final test form their mean score was 18.7. (The final mean scores for control *S*s were 18.9 on Form A and 18.6 on Form B.) This difference is not statistically significant and it is safe to assume that differences in mean scores of the experimental *S*s were due to the experimental variables rather than differences in the test forms.

These data indicate that the members of homogeneous groups can achieve significantly better decisions by solving their problems cooperatively than they can through voting or by individual effort. Majority decisions, when all deadlocks can be successfully resolved, can produce better results than are obtained from the averaging of individual efforts. But in three out of four of the conditions observed in this experiment, majority decisions proved to be no better than, or inferior to, the decisions of individual members of the same groups.

[2] In both of these cases the groups contained individuals who received almost perfect initial scores.

DISCUSSION

The results of the first phase of this experiment need to be interpreted in the light of early research on collective judgments. Whether they explained the finding on statistical or psychological grounds, Watson (10), Gordon (3), Stroop (8), and Gurnee (4) found grouped judgments superior to those of the average individual and equal to those of the superior individual working alone. This conclusion is not supported by our data. When deadlocks are resolved on the basis of statistical probabilities, majority decisions are found to be no better than those of the average member of homogeneous groups.

The explanation for the difference in results seems to lie partly in the character of the tasks and partly in the methods of grouping data. Some of the problems used by these investigators involve what may be called additive activities. Whenever individual efforts are additive or cumulative, the larger the group the greater should be the advantage from combining individual data. Testing the accuracy of conclusions drawn from given arguments is not the same kind of problem. One answer simply cannot be added to another. A second explanation for the difference is found in the manner of grouping individual decisions. The pooling of data in previous studies combined the heterogeneous opinions of 10 to a 100 individuals. In averaging data the greater the number and range of scores, the larger the gain from cancelling out individual errors. In this case, only four to six opinions from individuals of comparable ability determined the

decision. Majority rule may prove a convenient political device for averaging individual preferences; but our results suggest that in small, homogeneous groups or committees, majority rule, when it precludes discussion or debate, is likely to be less effective than the personal judgment of superior members of the group.

After discussion, however, experimental groups produced decisions that were far superior to those of members working alone or through majority rule. Moreover, group decisions on the test problems were reached within the same period of time allotted to individuals.

Several hypotheses are offered in the literature for the high quality of group judgments. Watson found group decisions superior because of the influence of the ablest member.

> In measuring the output of a group, either when working along cooperative group-thinking lines or when the project permits the simple compilation of individual efforts, it matters little about the ability of the poorest or even average member of the group. The results seem to show primarily what the few ablest in the group have produced (10, pp. 333–334).

This hypothesis, though generally tenable, seems inadequate to explain the results of this experiment. Groups were made up of students whose initial performance indicated a common aptitude for selecting logical conclusions from given arguments. The grading of items on the "Recognition of Valid Conclusions" test is such that persons who get similar scores are likely to possess not only the same level of ability but similar habits of thinking.

Another theory, suggested by Gurnee (4) and Thorndike (9), is that the better performance of the group is due to the social influence of the more confident group members who are more often right than wrong. It is difficult to see how this factor could have played a large part in the results. It would seem likely that students with similar patterns of right and wrong answers would share somewhat similar patterns of confidence about their answers. If so, this factor can be minimized.

It is necessary to go beyond these hypotheses to explain how correct solutions were reached by groups whose members made similar or identical errors when working alone. The diagnostic discussions and the analysis of recorded group sessions furnish additional clues to the psychological factors affecting the high level of group performance.

Membership in the experimental groups produced a higher level of interest in the successful completion of the task. Ss concentrated more intently on the assigned problems after being appointed to a group than they did when solving the problems individually. Group members found themselves more and more deeply involved as they proposed, and were forced to defend, their ideas. Participants identified with their own groups to such a degree that when some members became fatigued, others urged them to continue working.

Membership in the experimental groups had an inhibiting as well as facilitating effect. Knowledge that one's opinions were to be shared publicly made group members more cautious and deliberate in their own thinking. The necessity of explaining a conclusion forced many students to be more self-critical. Errors that might have been committed privately were checked before they were communicated to others.

Groups had greater critical resources than did individuals working alone. In spite of the uniform level of ability, group members saw different issues and a larger number of issues than a single person did working alone. A greater number of viewpoints increased the group's chances of selecting a valid one. Even the poorest members contributed significantly to the quality of the group product. Remarks that went no deeper than "I don't understand" or "That's absurd" often saved the group from error by forcing others to justify their opinions and in so doing disprove their own conclusions.

A more objective view of the problem resulted from competition between the private prejudices of group members. The test arguments were stated in loaded terms designed to make the choices between conclusions as difficult as possible. Each individual, however, brought a different set of values to his group. When arguments were stated so they appealed to persons of one persuasion, those in opposition were anxious to detect their error. In this way, liberals counteracted conservatives, Republicans offset Democrats, and "independents" guarded against critical lapses on the part of fraternity members. Groups were forced to

become more objective, and this, of course, increased their chances of drawing valid conclusions. The significance of this one factor alone would be hard to overestimate.

Discussion of the test items also prevented other incidental mistakes from occurring. Some groups had to check their instructions several times because members had different interpretations of them. Discussion often led to a clarification of terms used in the test, and, where logical fallacies spring from ambiguous terms, this may account for some of the gains. A number of groups formulated general principles as they went along to help them avoid repeating errors in later problems.

What, then, prevented experimental groups from attaining even higher scores than they did? Analysis of the transcripts revealed two factors that together accounted for a majority of the group errors. The first was that group members agreed immediately and unanimously upon the wrong answer to a problem. Further study of the issue was then considered unnecessary and wasteful. This is the same factor that Jenness, following F. H. Allport, refers to as the "impression of universality" (5). Agreement becomes the criterion of correctness. Maier (7) suggests that provoking arguments under these circumstances leads to better judgments. The virtue of disagreement and the possible function of a "No-Man" in group deliberations, needs further testing.

The second factor was that groups, when they reached a deadlock, were unable to use their differences of opinion for their own advantage. When conflicts became intense they were resolved by surrender of the less aggressive members or by compromising on a third solution which was almost always incorrect but served to protect the egos of the parties to the controversy. Apparently disagreement stimulates thought up to a point; beyond that point, groups may lack the patience and skill to exploit it.

Discussion, as a preliminary to group decisions, causes groups to examine a problem more thoroughly and to consider a wider number of solutions. It encourages individuals to think more carefully and in sharing opinions to expose the logic of their position to the inspection of others. Membership in a group produces a sense of responsibility which intensifies and sustains effort. The biasing effect of private prejudice may be counteracted leading to a more objective view of the issues. The data of this study indicate that the answer to the question of whether group opinion is better than individual opinion because of the influence of the superior person or because of the discussion process itself is that discussion inherently contains psychological pressures and motivations which, if not abused, tend to produce superior judgments on complex intellectual problems. Individual decisions and collective judgments lack the additional ingredient supplied by interaction which permits a group to outperform its own members.

SUMMARY

The performance of individuals working alone, under majority rule, and as members of discussion groups were compared on a complex intellectual task. Individual judgment was measured by administering a test of ability to draw logical conclusions from given arguments. Individuals receiving similar scores were assigned to the same experimental groups so that the factor of distributed ability would be reduced to a minimum. The votes of members of the homogeneous groups were mathematically tallied to determine the results under conditions of majority rule. A second form of the test was completed as a group undertaking and the scores compared with individual and majority scores. The results indicated that:

1. Majority decisions, when deadlocks are evenly divided between right and wrong answers, are not significantly different from those made by the average individual and are inferior to those of the best member of the group working alone.

2. Group decisions, reached through cooperative deliberation, are significantly superior to decisions made by individual members working alone and to majority rule.

The superiority of group judgments was found not to be a reflection of the wisdom of the superior member of the group but a result of psychological factors inherent in discussion. Participation in a group led to more serious concentration on the task and to more enthusiastic individual effort. Group discussion was found to stimulate more careful thinking, to lead to a consideration of a wider range of

ideas, and to provoke more objective and critical testing of conclusions.

REFERENCES

1. BRADLEY, E. E. Formal validity in problem solving. Unpublished doctoral dissertation, Northwestern Univer., 1950.
2. DASHIELL, J. F. Experimental studies of the influence of social situations on the behavior of human adults. In C. Murchison (Ed.), *Handbook of social psychology*. Worcester: Clark Univer. Press, 1935. Pp. 1097–1158.
3. GORDON, K. H. Group judgments in the field of lifted weights. *J. exp. Psychol.*, 1924, **7**, 398–400.
4. GURNEE, H. A comparison of collective and individual judgments of fact. *J. exp. Psychol.*, 1937, **21**, 106–112.
5. JENNESS, A. The role of discussion in changing opinion regarding a matter of fact. *J. abnorm. soc. Psychol.*, 1932, **27**, 279–296.
6. KELLEY, H. H., & THIBAUT, J. W. Experimental studies of group problem solving and process. In G. Lindzey (Ed.), *Handbook of social psychology*. Cambridge: Addison-Wesley, 1954. Pp. 735–785.
7. MAIER, N. R. F., & SOLEM, A. R. The contribution of a discussion leader to the quality of group thinking. *Hum. Relat.*, 1952, **5**, 277–288.
8. STROOP, J. R. Is the judgment of the group better than that of the average member of the group? *J. exp. Psychol.*, 1932, **15**, 550–562.
9. THORNDIKE, R. L. The effect of discussion upon the correctness of group decisions, when the factor of majority influence is allowed for. *J. soc. Psychol.*, 1938, **9**, 343–362.
10. WATSON, G. B. Do groups think more efficiently than individuals? *J. abnorm. soc. Psychol.*, 1928, **23**, 328–336.

Received September 12, 1957.