MANAGEMENT SCIENCE
Vol. 28, No. 10, October 1982
Printed in U.S.A.

GROUP CONFIDENCE PRESSURES IN ITERATIVE DECISIONS*

DAVID M. BOJE† AND J. KEITH MURNIGHAN‡

This study investigated the effects of two group decision making techniques on a set of four problems in groups of sizes 3, 7, and 11. Participants included 192 male and 132 female undergraduates. Estimates that could be evaluated for accuracy for each of the problems were collected for a series of three trials. One set of groups received face-to-face verbal feedback from each other, while the other set received written feedback. These data were compared to mean estimates obtained from randomly selected, pooled individual estimates. The results suggested that the pooled individual estimates were somewhat more accurate than those obtained from either of the interacting groups. At the same time, all individuals became more confident of their answers, suggesting the possibility of groupthink. No effects for different group sizes were found, possibly due to the constraints imposed by the structured nature of the two techniques.

(GROUPS; DECISION MAKING; DELPHI; CONFIDENCE; ACCURACY)

1. Introduction

Structured approaches to group decision-making in organizations have been suggested recently (e.g., [10]) as ways to improve effectiveness and efficiency. Structuring out some of the opportunities for interpersonal interaction (1) may avoid the "process" losses often noted in groups [24], and (2) may reduce the possibility of groupthink [15], where individuals become oriented to the interpersonal climate within the group rather than challenging the accuracy of their decisions. Early research on group dynamics (e.g., [4]) also indicates that more informal (less structured) group meetings subject group members to strong social pressures that may inhibit creativity and accuracy in problem solving efforts. The present study was designed to evaluate the effects of two structured approaches to group decision making on the confidence of group members and the accuracy of their decisions. As is typical in the group problem solving literature, the results were also compared to the solutions of a set of non-interacting individuals to assess whether controlled group interaction led to improved decision accuracy and improved confidence in the decisions reached.

The orientation towards decision making taken here is that organizational participants in many settings find themselves embedded in an ambiguous and chaotic stream of persons, problems and solutions [17]. As such, participants spend a great deal of time and energy attempting to verify and rationalize the "accurate" matching of problems with appropriate solutions. Group decisions not only involve accurate and efficient decision making, they also involved the process of (1) self justification that a decision is accurate; (2) group rationalization that it is accurate; and (3) explaining or legitimizing the group decision to the powers that be.

Groups often engage in problem solving interactions and, even if they do not produce more accurate decisions, they may become more confident that what they are deciding to do is appropriate and correct. The two techniques to be explored here were designed in part to maximize this effect, what we term "group confidence pressure." We will begin by briefly describing the respective techniques.

*Accepted by Arie Y. Lewin; received May 20, 1980. This paper has been with the authors 4 months for 2 revisions.

†University of California, Los Angeles.
‡University of Illinois.

0025-1909/82/2810/1187$01.25

the Delphi process [7]. Increased accuracy occurred over iterations on the same problem; indeed, several iterations are an integral part of the Delphi technique. One might expect (although there is no evidence to date on the issue) that the "face-to-face feedback" procedure might also result in increased accuracy over iterations. Thus, each group made three estimates for each of the problems, receiving feedback for previous estimates. The comparison set of randomly pooled individual estimates did not benefit from feedback; three iterations were used in this condition, however, to increase comparability.

This research, then, attempted to determine the differential confidence and accuracy of "written feedback" and "face-to-face feedback" procedures on four previously researched problems. In addition, the study focused on the effects of groups of different sizes and three iterations of the techniques with variations in the structure of the group and/or its process to further clarify potential interactive effects.

2. Method

Subjects

Participants included 192 male and 132 female undergraduate students enrolled in an introductory organizational behavior course at a large midwestern university. Each received credit toward a course requirement for participation in the study. Most groups were a mix of males and females.

Procedure

Subjects were told that they would be using one of several decision making techniques analogous to those used by groups in organizations. Each individual was given the four problems; all were encouraged to try to be as accurate as possible in their estimates. In the "face-to-face feedback" (FF) groups, subjects were introduced to each other, were seated around a table, began work on the problems individually, maintained silence as they generated their answers, and discussed information after each listing of solutions. As in the "written feedback" (WF) conditions, group members were asked for an estimate of the correct answer and one fact or reason in support of the estimate, without discussion. In the FF groups, individuals presented their estimates in round-robin fashion to the experimenter, who recorded them on the blackboard. After a brief discussion of the estimates, primarily directed to the group leader (the experimenter), the group members were asked to make their second individual estimate. The round-robin was again followed by discussion for the second and third trials. Unlike the Nominal process proposed by Delbecq and Van de Ven [9], the estimates were not put to a vote. As in the other conditions, the FF group's estimates were operationally defined as the mean of the individual responses at each trial.

In the WF and individual conditions, subjects were seated behind opaque partitions, restricting interpersonal contact. In the WF condition, each person's individual estimate and reason or fact was copied verbatim and distributed to the other group members as quickly as possible. In the individual condition, subjects were told to reflect upon the problems and to think of additional information that might be important in finding the solution to the problems. In each case, the group estimate was, again, the mean of the individual responses at each trial.

Before debriefing, each subject was asked to complete a brief questionnaire concerning his/her reactions to the decision technique and the experimental task.

Problems

Two subjective likelihood problems were adapted from Gustafson, *et al.* [11]; two

problems were also taken from the *1976 World Almanac*. The two subjective likelihood problems (weight and height) and the two Almanac problems (Jupiter and dollars) were consistent with those employed in previous research in this area.

(1) *Weight*: The average weight of men is 154 pounds (69.9 kg). The average weight of women is 128 pounds (58.1 kg). Out of a random sample of 100 people, all of whom are 150 pounds (68.0 kg) in weight, how many would be male? (Correct answer: 81.13);

(2) *Height*: The average height of men is five feet nine inches (1.75 m). The average height of women is five feet four inches (1.63 m). Out of a random sample of 100 people, all of whom are 68 inches tall (1.73 m), how many would be male? (Correct answer: 64.29);

(3) *Jupiter*: The earth's moon has a diameter of 2,160 miles (3,476 km). The diameter of the sun is 864,000 miles (1,390,435 km). What is the diameter of the planet Jupiter at its equator? (Correct answer: 86,000 miles; 139,687 km); and

(4) *Dollars*: Dollar bills measure 2-5/8″ (6.67 cm) by 6 = 1/8″ (15.56 cm) with a thickness of 0.0043″ (0.109 mm). New notes will stack 233 to an inch, if not compressed. How many dollar bills would be needed to weigh exactly one pound (0.45 kg)? (Correct answer: 490).

Following Phillips and Edwards [21], subjects responded to the subjective likelihood problems on a logarithmically calibrated scale of odds to reduce the potential conservatism effect. To control for possible order effects, problems were arranged in seven separate random orders; each order was used at most once within each treatment condition.

In addition to a numerical estimate and a fact or reason for each problem, subjects indicated the confidence they had in each of their answers on seven-point scales.

Design

Three levels of group size (3, 7, and 11), three types of decision procedures (WF, FF and Individual), four problems (weight, height, Jupiter and dollars), and three trials were examined in a $3 \times 3 \times 4 \times 3$ design. Problems and trials were repeated measures; group size and procedures were between factors.

The dependent variables were: (1) The confidence rating reported by individuals in each trial and for each problem; (2) a measure of group accuracy (the deviation of the group mean from the correct answer, standardized to allow comparisons among problems); and (3) individual responses on ten questionnaire items (see Table 3 for a partial listing of the questions) on the efficiency of the technique they used.

Due to an insufficient number of subjects, there were a maximum of seven groups in each of the procedure/size conditions. In the WF condition, there were 7 three-person, 6 seven-person, and 6 eleven-person groups for a total of 129 participants. In the FF conditions, there were 7, 7, and 5 groups, respectively, and 125 participants. And in the individual conditions, there were 7, 7, and 6 "groups," randomly selected with replacement from a pool of 70 participants.

3. Results

The mean estimates for each problem by the groups using different procedures are shown, for each of the three trials, in Table 1. Separate results are not depicted for the different group sizes because size yielded no significant effects on the accuracy of the group's estimates.

The first analysis was a multivariate analysis of variance with procedures (3), size (3), problems (4), and trials (3) as independent variables and confidence and accuracy as the dependent measures. Significant main effects were found for procedures

TABLE 1

Mean Estimates for Each Problem by Each Procedure Over the Three Trials

Problem (Correct Answer)	Procedure	Trials 1	2	3
Weight (81.13)	Face-to-Face Feedback	73.75	74.22	74.45
	Written Feedback	73.96	76.74	78.64
	Individual (No Feedback)	73.21	69.53	72.97
Height (64.29)	Face-to-Face Feedback	70.70	76.79	77.39
	Written Feedback	65.42	69.94	72.14
	Individual (No Feedback)	67.90	69.41	70.51
Jupiter (86,800)	Face-to-Face Feedback	154,000	191,400	198,600
	Written Feedback	95,900	155,500	173,100
	Individual (No Feedback)	119,600	140,400	136,600
Dollars (490)	Face-to-Face Feedback	969.4	1081.3	984.2
	Written Feedback	653.1	877.1	816.1
	Individual (No Feedback)	942.1	973.2	1017.6

[multivariate $F(4,96) = 4.32$, $p < 0.003$], problems [multivariate $F(6,292) = 70.97$, $p < 0.0001$], and trials [multivariate $F(4,194) = 24.16$, $p < 0.0001$]. Significant interactions were also found between procedures and trials [multivariate $F(8,194) = 2.24$, $p < 0.03$], problems and trials [multivariate $F(12,586) = 9.57$, $p < 0.0001$], and procedures, problems, and trials [multivariate $F(24,586) = 1.62$, $p < 0.054$].

The corresponding univariate analyses of variance suggest that the participants' confidence ratings contributed more than accuracy to these significant effects. The main effect for trials [$F(2,98) = 61.76$, $p < 0.0001$] indicates that confidence increased over trials. The main effect for problems [$F(3,137) = 245.80$, $p < 0.0001$] indicates that participants' confidence was greater in the subjective likelihood problems (weight and height) than the Almanac problems (Jupiter and dollars). Respondents consistently reported the latter as being more difficult to decide. The main effect for procedures [$F(2,49) = 6.74$, $p < 0.003$] indicates that participants' confidence was stronger when they used either the WF or FF procedures.

Due to the large number of means involved in the procedures by problems by trials interaction [$F(12,294) = 2.21$, $p < 0.02$] it is not clear which means contributed most to the effect. The problems by trials interactions [$F(12,294) = 21.15$, $p < 0.0001$], however, appears to be due to considerably greater increases in confidence for the Almanac problems, especially for the dollars problem, than for the probability estimates. This in turn may be explained by the observation that participants conveyed more written and verbal information on these problems. This finding suggests that

TABLE 2

Mean Accuracy Scores[a] for the Procedures × Trials Interaction

Procedures	Trials 1	2	3	Mean
Face-to-Face Feedback	0.163	0.270	0.180	0.205
Written Feedback	− 0.186	− 0.054	− 0.015	− 0.085
Individiual (No Feedback)	0.020	− 0.198	− 0.158	− 0.112
Mean	− 0.001	− 0.006	0.003	—

[a]The correct answers for the four problems, after being transformed to standardized scores, differed; the mean of the accurate scores, transformed, was −0.430.

TABLE 3

Mean Response and Summary of Analyses for the Questionnaire Items[a]

	Written Feedback	Face-to-Face Feedback	Individual (No Feedback)	$F(2,49) =$	$p <$
1. How would you rate the procedure you used as a way of making decisions? (Very Poor—Excellent)	2.88_{ab}	3.32_{a}	2.69_{b}	3.38	0.05
2. How much did you enjoy working on this project? (Not at All—A Great Deal)	2.64_{b}	3.54_{a}	2.56_{b}	16.12	0.001
3. To what extent do you feel this decision technique was appropriate for these problems? (Not at All—A Great Deal)	2.84_{ab}	3.33_{a}	2.69_{b}	4.78	0.01
4. Do you feel you had enough time to answer these questions? (Too Little—Too Much)	4.64_{b}	4.74_{b}	5.55_{a}	19.67	0.001
5. Do you feel greater accuracy would result if *more* people worked with you on this project? (Less Accurate—More Accurate)	5.23_{ab}	4.94_{b}	5.69_{a}	7.30	0.01
6. To what extent did you feel free to contribute your ideas? (Not at All—A Great Deal)	4.44_{b}	5.40_{a}	4.09_{b}	21.75	0.001
7. Do you feel others working on this problem influenced your response? (Not at All—A Great Deal)	3.85_{b}	4.72_{a}	1.44_{c}	86.12	0.001

[a] The multivariate F-ratio for procedures was 10.99, $p < 0.0001$, with 20 and 80 df.

Means sharing a common subscript within the levels of each main effect are *not* significantly different from one another at the 0.05 level using the Newman–Keuls procedure.

Almanac problems may be more ambiguous than the probability estimates, thus leading to greater information processing [23].

The only significant univariate analysis for accuracy was the procedures by trials interaction, $F(4,98) = 3.26$, $p < 0.02$. The means for this effect are shown in Table 2. While the WF and FF procedures became less accurate over trials, the individual procedure became more accurate over trials. Although the associated confidence ratings did not reach standard significance levels for this interaction, $F(4,98) = 2.03$, $p < 0.10$, the results suggest, as do the main effect for trials, that individuals' confidence increased over trials, even as accuracy decreased. Also, although the main effect for procedures was not significant for accuracy [$F(2,49) = 2.25$, $p < 0.12$], the figures suggest that the FF procedure led to greater overestimates than the other two techniques. Overall, the results discount the advisability of multiple trials for problems like these: One's first guess may be best.

Because the questionnaire items were not repeated measures, they were included as dependent measures in a multivariate analysis of variance with procedures and size as independent variables. Both main effects were significant: multivariate $F(20,80) = 10.99$, $p < 0.0001$ and multivariate $F(20,80) = 1.75$, $p < 0.05$, respectively. The size effect was primarily determined by two items: the three-person groups felt more people would increase their group's accuracy, especially compared to the eleven-person groups, $F(2,49) = 4.44$, $p < 0.02$, and people in the eleven-person groups felt significantly less free to contribute their ideas than people in the seven-person groups, $F(2,49) = 3.65$, $p < 0.05$. It is important to note here that group size had no significant effects on accuracy or confidence.

Table 3 displays the means and F-ratios for the significant univariate effects for procedures. The results suggest that the WF procedure is only somewhat superior in the eyes of the participants than working alone, and that the FF procedure generally resulted in the most positive perceptions of effectiveness, satisfaction, and fredom. These effects support earlier findings [26] that indicated Nominal group members felt greater satisfaction than Delphi group members or freely interacting groups.

4. Discussion and Conclusions

The most notable and consistent findings in this study were for the confidence individuals expressed in their estimates. High confidence ratings were observed in the FF and WF groups, on the subjective likelihood problems, and as the trials increased. Increases in confidence over trials were observed for the Almanac problems, where individuals in the second and third trials had the opportunity to express some expertise on the problems (e.g., "I worked in a bank last summer," or "I took a course in astronomy."). While a rational model might expect that confidence would increase as an individual's estimates became more accurate, the reverse occurred in this study. Confidence increased, at least in the FF and WF groups, as accuracy dropped. Possible explanations include a desire to be confident when interacting with other people or as one "practices" and has greater experience with a problem. These data clearly indicate that factors other than accuracy may contribute to feelings of confidence, especially when no feedback is provided.

Increases in confidence paired with reduced accuracy also suggest that a form of groupthink [15] may have been operating in these groups. Greater influence may have been exerted by group members espousing the least accurate estimates. Indeed, the "critical task contingencies" [13] for these problems may center around the availability of accurate information in each group. In the absence of relevant information (which may not have been unusual given this set of problems), it might not be surprising that a non-interacting process yields the most accurate answer. These results contradict

conventional wisdom in employing structured decision techniques in such ambiguous and imperfect information areas as forecasting and strategy formation.

Similarly paradoxical is the relationship between the accuracy scores and subjects' perceptions of the procedures. The FF procedure yielded the least accurate answers, but the most positive affective responses. It seems obvious that the interpersonal interaction within the group led to more positive ratings on the questionnaire items, and that isolation or working on the problems alone is much less pleasant. The "misperceptions" by the FF group members further suggest the presence of "group confidence pressures."

Group confidence pressures, as we conceptualize them, depend on both verbal and nonverbal feedback from members of one's group, and result in both greater confidence and increased satisfaction ratings. Seeing and hearing other people express their opinions, and having to express your own opinions in their presence, contributes to feelings of compliance, confidence, and satisfaction. All three responses are produced by the knowledge that others have observed your behavior. Each individual justifies his/her position by being more confident and satisfied with each succeeding public statement. Thus, group confidence pressures can be interpreted within the more encompassing frameworks of cognitive dissonance and self-perception theories [1], [3].

In terms of effectiveness then, these data suggest that for problems where accuracy is important, group members should not interact with one another or even exchange information. In situations where accuracy is less important (accuracy may not even be a relevant criterion for many problems), and satisfaction is more important, interaction appears to be desirable. Future research might test these notions and pursue the task and group process characteristics that yield positive affect and/or accuracy.

Even though group size has often been cited as a critical variable in the study of groups [6], [8], size in this study affected neither accuracy nor confidence. The fact that size did affect perceived freedom in contributing one's ideas replicates earlier results [e.g., 4]. Also somewhat expected were the results indicating that the small groups felt additional members would increase their group's accuracy. The fact that size increases, however, did not increase group accuracy suggests that in groups that use *structured* decision processes, like those employed in this study, size has little opportunity to affect the outcomes of the group process. With less structured conditions, size may be more effectual. Other studies, possibly using additional sets of problems and less structured decision procedures might test this hypothesis.

This study was somewhat unusual in its use of more than a single problem to test the differences between two group decision procedures. But even using four different problems, which spanned two types of previous research studies, further information is necessary to adequately explore the effectiveness of different decision procedures. Recent papers by Guzzo [12] and Murnighan [19] suggest strategies to be considered for different problem types, but little empirical evidence exists which spans, for instance, problems where accuracy is an issue versus problems where accuracy is not even relevant. While this study focused on the accuracy issue, and how it interacted with different problems, different feedback procedures, and different group sizes, the findings of Van de Ven [26], where accuracy was not an issue, stand in marked contrast. The morale of the story appears to be that the dynamics of group decision making are particularly complex, and that large scale research projects may be necessary to understand the intricacies of procedural, task, and group member interactions.

References

1. Aronson, E., "The Theory of Cognitive Dissonance: A Current Perspective," in L. Berkowitz (ed.), *Advances in Experimental Psychology*, Vol. 4, Academic Press, New York, 1969, pp. 1–34.

2. Asch, S. E., "Opinions and Social Pressure," *Scientific American* (November 1955).
3. Bem, D. J., "Self-Perception: An Alternative Interpretation of Cognitive Dissonance Phenomena," *Psychological Rev.*, Vol. 74 (1967), pp. 183–200.
4. Cartwright, D. and Zander, A., *Group Dynamics*, Harper and Row, New York, 1968.
5. Cottrell, N. B., "Social Facilitation," in C. G. McClintock (ed.), *Experimental Social Psychology*, Holt, Rinehart, and Winston, New York, 1972.
6. Cummings, L. L., Huber, G. P. and Arendt, E., "Effects of Size and Spatial Arrangements on Group Decision Making," *Acad. Management J.*, Vol. 17 (1974), pp. 460–475.
7. Dalkey, N. C., *Studies in the Quality of Life: Delphi and Decision Making*, Lexington Books, Lexington, Mass., 1972.
8. Davis, J. H., Laughlin, P. and Komorita, S. S., "Small Group Research in Social Psychology, 1968–1975," *Annual Rev. Psychology*, Vol. 24 (1976), pp. 1101–1188.
9. Delbecq, A. L. and Van de Ven, A. H., "A Group Process Model for Problem Identification and Program Planning," *J. Appl. Behavioral Sci.*, Vol. 7 (1971), p. 4.
10. ———, ———, and Gustafson, D. H., *Group Techniques for Program Planning: A Guide to Nominal and Delphi Processes*, Scott Foresman, Glenview, Ill., 1975.
11. Gustafson, D. H., Shukla, R. K., Delbecq, A. and Walster, G. W., "A Comparative Study of Differences in Subjective Likelihood Estimates Made by Individuals, Interacting Groups, Delphi Groups, and Nominal Groups," *Organizational Behavior and Human Performance*, Vol. 9 (1973), pp. 280–291.
12. Guzzo, R. A., *The Decision Making Quality of Managerial Groups*, Unpublished dissertation, Yale University, 1979.
13. Hackman, J. R. and Morris, C. G., "Group Tasks, Group Interaction Process, and Group Performance Effectiveness: A Review and Proposed Integration," in L. Berkowitz (ed.), *Advances in Experimental Social Psychology*, Vol. IX, Academic Press, New York, 1975.
14. ———, and Vidmar, N., "Effects of Size and Task Type on Group Performance and Member Reactions," *Sociometry*, Vol. 33 (1970), pp. 37–54.
15. Janis, K., *Victims of Groupthink*, Houghton Mifflin, Boston, Mass., 1972.
16. Maier, N. R. F., *Principles of Human Relations*, Wiley, New York, 1952.
17. March, James G. and Olsen, J. P., *Ambiguity and Choice in Organizations*, Universitetsfarlaget, Bergen, Norway, 1976.
18. Miner, F. C., "A Comparative Analysis of Three Diverse Decision Making Approaches," *Acad. Management J.*, Vol. 22 (1979), pp. 81–93.
19. Murnighan, J. K., "Group Decision Making: What Strategies Should You Use?," *Management Rev.*, Vol. 70, No. 2 (February 1981), pp. 55–62.
20. Pfeiffer, J., *New Look at Education*, Odyssey Press, New York, 1968.
21. Phillips, L. D. and Edwards, W., "Conservatism in a Simple Probability Inference Task," *J. Experimental Psychology*, Vol. 72 (1966), pp. 346–354.
22. Salancik, G. R., "Commitment and the Control of Organizational Behavior and Belief," in B. M. Staw and G. R. Salancik (eds.), *New Directions in Organizational Behavior*, St. Clair Press, Chicago, Ill., 1977.
23. Shaw, Marvin, E., *Group Dynamics: The Psychology of Small Group Behavior*, McGraw-Hill, New York, 1971.
24. Steiner, E., *Group Process and Productivity*, Academic Press, New York, 1972.
25. Van de Ven, A. and Delbecq, A. L., "Nominal Versus Interacting Group Processes for Committee Decision Making," *Acad. Management J.*, Vol. 14 (1971), pp. 203–213.
26. ———, *Group Decision Making and Effectiveness: An Experimental Study*, Kent State University, Kent, Ohio, 1974.