COMPUTERIZED CONFERENCING & COMMUNICATIONS CENTER

at

NEW JERSEY INSTITUTE OF TECHNOLOGY

THE EFFECTS OF FORMAL HUMAN LEADERSHIP AND COMPUTER-GENERATED DECISION AIDS ON PROBLEM SOLVING VIA COMPUTER: A CONTROLLED EXPERIMENT

By

Starr Roxanne Hiltz, Kenneth Johnson, and Murray Turoff

Research Report Number 18

Computerized Conferencing and Communications Center

c/o Computer & Information Science Department
New Jersey Institute of Technology
Newark, N. J. 07102

THE EFFECTS OF FORMAL HUMAN LEADERSHIP AND COMPUTER-GENERATED DECISION AIDS ON GROUP PROBLEM SOLVING VIA COMPUTER: A CONTROLLED EXPERIMENT

Starr Roxanne Hiltz, Kenneth Johnson, and Murray Turoff

Research Report Number 18

Computerized Conferencing and Communications Center, NJIT

December, 1982

The research reported here was conducted with a grant from the Division of Mathematical and Computer Sciences, National Science Foundation (MCS-78-ØØ519). The opinions and findings are solely those of the authors and do not necessarily reflect the views of the National Science Foundation.

We are grateful to Nancy Rabke for her assistance in all phases of this project. Tom Moulton and James Whitescarver deserve credit for the many long nights and days which they spent designing and endlessly testing the software for total computer administration of the experiment, and for standing by online in case of difficulty during the experimental runs. Charles Aronovitch made many helpful contributions to the design and testing phase of the experiment. Elaine Kerr contributed a careful critical reading of the manuscript. Andrew Finn contributed a helpful cross-checking of the accuracy of the coded data files. Anita Graziano supervised the production of this final report. We would also like to thank the organizations which took part in the study; they contributed travel funds as well as their time. We hope that they may have learned something in the process, too.

ABSTRACT

Twenty-four groups of five professionals and managers within a variety of organizations were given the task of using a computer conference to reach agreement on the best solution to a ranking problem.

The independent variable is the structure of the conferencing capability used. Two alternative means of structuring the conferences were employed, in a two-by-two factorial design. Groups with "Human Leadership" elected one of their members to lead the group in its decision making discussion. Groups with "Computer Feedback" were given periodic tables which displayed the current "group decision" in terms of the mean rankings of items, and the degree of consensus about each of these items.

Dependent variables include:

- .Quality of decision
- .Degree of consensus
- .Amount of discussion and reranking activity
- .Equality of participation
- .Subjective satisfaction

Covariates include initial (pre-discussion) quality of decision, typing speed, knowledgability of the leader, age, and sex.

For this experiment, with small groups, human leadership was more effective than computer feedback for improving consensus and quality of decision.

TABLE OF CONTENTS

CHAPTER ONE
AN OVERVIEW OF THE EXPERIMENT

INTRODUCTION.....1
Computer-Mediated Communication: Generalizations and Variations.....3
Background: The Prior Experiment.....5
EXPERIMENTAL DESIGN.....8
The Independent Variables: Structuring the Group Process.....9
Dependent and Process Variables.....15
Hypotheses.....17
Covariates.....18
SUBJECTS AND PROCEDURE.....19
Description of the Analysis of Variance Designs.....22
SUMMARY.....23

CHAPTER TWO
QUALITY OF DECISION

MEASURES OF QUALITY OF DECISION.....24
The Decision Data.....24
The Percentage Improvement Measure.....25
"Collective Intelligence".....26
DIFFERENCES IN QUALITY OF THE GROUP DECISION.....28
The Selection and Performance of Leaders.....32
Influence of the "Best" Member.....33
The Effect of Group Composition.....34
"COLLECTIVE INTELLIGENCE," BY CONDITION.....36
SUMMARY.....39

CHAPTER THREE
ABILITY TO REACH CONSENSUS

REACHING A GROUP DECISION.....41
RESULTS FOR FINAL INDIVIDUAL RANKINGS.....46
FACTORS RELATED TO THE ABILITY TO REACH CONSENSUS.....48
SUMMARY.....50

CHAPTER FOUR
VARIATIONS IN GROUP PROCESS RELATED TO STRUCTURE AND SUBJECT CHARACTERISTICS

DOMINANCE.....52
THE EFFECTS OF HUMAN LEADERSHIP AND FEEDBACK ON AMOUNT OF DISCUSSION AND RANKING.....54
THE EFFECTS OF SUBJECT CHARACTERISTICS ON PERFORMANCE.....57
PROCESS VS. OUTCOME.....60
SUMMARY.....61

CHAPTER FIVE
SUBJECTIVE SATISFACTION

EFFECTS OF EXPERIMENTAL CONDITION ON SUBJECTIVE SATISFACTION....64
VARIATIONS BY SUBJECT CHARACTERISTICS....67
The Effect of Age....67
Sex and Subjective Satisfaction....69
Typing and Subjective Satisfaction....70
Effect of Previous Computer Terminal Experience....70
GROUP DIFFERENCES....71
SUMMARY....72

CHAPTER SIX
METHODOLOGY: THE AUTOMATED EXPERIMENT

COMPUTER AND HUMAN ROLES IN CONDUCTING THE EXPERIMENT....74
Taking the Experiment Into the Field....78
Training and Monitoring Aids....79
Problems: Automated Errors....81
Automated Analysis....82
A SIMILAR AUTOMATED EXPERIMENT....83
CONCLUSION....86

CHAPTER SEVEN
SUMMARY AND CONCLUSIONS

SUMMARY OF FINDINGS....87
Decision Quality and Degree of Consensus....87
Group Process....89
Subjective Satisfaction....90
Variations Associated with Subject Characteristics....91
The Pervasive Influence of Group Differences....92
NOTES ON STRUCTURE....93
IMPLICATIONS FOR DECISION SUPPORT SYSTEMS....97

DDSS and the structure of Organizations: Some assertions....98
CONCLUSIONS....100

APPENDICES....102
REFERENCES....116

LIST OF TABLES

TABLE PAGE

1-1 Text-Only Table Received in All Conditions (Example for the Practice Problem).....13
1-2 Sample of Computer Feedback Table (for the Practice Problem).....13
1-3 Design of the Experiment.....14
2-1 Mean Initial Deviation Scores by Condition.....27
2-2 Mean Group (Post Discussion) Deviations from Correct Answer, by Condition.....30
2-3 Percentage Improvement in Deviation from Criterion, By Condition.....31
2-4 "Collective Intelligence," by Condition.....38
3-1 Initial (pre-discussion) Agreement, by Condition.....44
3-2 Degree of Consensus on Final Group Ranking.....45
3-3 Degree of Consensus on Final Individual Ranking.....47
4-1 Variations in Number of Comments, by Condition.....56
4-2 Mean Number of Run Re-Rankings, by Condition.....57
4-3 Significant Pearson Correlation Coefficients Between Subject Characteristics and Performance Variables.....59
5-1 Perceived Friendliness of the Group, by Condition.....66
5-2 Perception of Having Reached Consensus, by Condition.....66
5-3 Correlations Between Age and Subjective Satisfaction with CC.....68
5-4 Satisfaction with One's Performance, by Age.....68
5-5 Sex by Satisfaction with CC for Getting to Know Someone...69
5-6 Previous Computer Terminal Experience by Satisfaction with One's Performance.....71

CHAPTER ONE

AN OVERVIEW OF THE EXPERIMENT

INTRODUCTION

How do variations in the structure of computerized conferences affect the process and outcome of group problem-solving discussions? Is it possible to create software which is more effective for group problem solving than free-form or unstructured conferences? Or do all forms of computer-mediated communication systems have similar effects on group discussions? This is a report on a controlled experiment designed to explore these questions.

Much of the early research on the social effects of computer-mediated communication systems (CMCS) involved attempts to reach generalizations about the impact of this new medium. For example, Johansen, Vallee, and Spangler (1979:180-181) summarize a number of studies with the statement that "computer conferencing promotes equality and flexibility of roles in the communication situation" by enhancing candor of opinions and by helping to bring about greater equality of participation. On the basis of early pilot studies comparing face-to-face and computerized conferences, Hiltz and Turoff (1978:124) conclude that more opinions tend to be requested and offered in computerized conferences, but that there is also less explicit reaction to the opinions and suggestions of others, whether

agreement or disagreement. In terms of organizational impacts, Uhlig, Farber, and Bair (1979:306) state that "collaboration of groups of persons, whether on a report or a complex decision, is accelerated by the speed of communication, including distribution and feedback." (See Kerr and Hiltz, 1982, for a summary of the generalizations which emerge from the findings of eighteen research and development projects related to CMCS).

The second generation, so to speak, of research on CMCS seeks a better understanding of the conditions under which the general tendencies of the medium are stronger, weaker, or totally absent. Some of this research focuses on the structure or facilities of the computer-mediated communications system itself. For instance, recent work at the Institute for the Future deals not with the general social effects of the PLANET system, but with the effects of adding three specific tools designed to support specific group tasks to the basic conferencing program: "graphical communication...communication focused on the running of computer programs through its program workspace, and communication focused on the creation and editing of a document" (Lipinski, Spang, and Tydeman, 1980:159).

Current work at the New Jersey Institute of Technology focuses on the development and evaluation of a variety of new capabilities for computer-mediated communication systems. The goal is to discover the interactions among task types, communications structures, and individual or group attributes that will allow the selection of optimal system designs and implementation strategies to match

variations in user group characteristics and types of tasks or applications. The research program involves a combination of field trials and controlled experiments. This report describes the second controlled experiment.

Computer-Mediated Communication: Generalizations and Variations

In computerized conferences, group members communicate by typing and reading on computer terminals rather than by speaking, listening, and exchanging nonverbal gestures. Each person types an entry without interruption and then receives any waiting communications. The communication channel is therefore missing many features of "normal" face-to-face communication, such as instantaneous receipt of communications and nonverbal cues (i.e., eye glance, facial expressions, tone of voice, and gestures). On the other hand, the presence of the computer in the communications loop provides some communication possibilities not available in a face-to-face meeting. For example, all participants can think as long as they want, without being interrupted by others, before making their comments. The participants can be on line at the same time in different locations ("synchronous" conferences or message exchanges), or more usually, sending and receiving communications at the time of their own choosing, with the computer storing waiting communications ("asynchronous" conferences). In synchronous exchanges, all can be typing at once, rather than having to take turns speaking. The printing-reading speed is faster than speaking-listening speed (30 characters per second was used in this experiment; 120 characters per

second is not uncommon). The repeat key and special characters on the keyboard can be used to easily create linguistic and graphic effects, such as the use of a whole line of exclamation points or question marks for emphasis (Carey, 1980).

The main varieties of computer-mediated communication are "messages" or "electronic mail," which store and forward discrete communications and may be thought of as replacing the internal memo, the letter, or the phone call; document and file transfer-systems such as NLS (now called Augment), which allow communication through the transfer of files; and "conferencing systems" which are oriented toward group communication by maintaining a transcript of a single-subject discussion for a whole group, and by providing features such as voting and markers which indicate the location of each participant in the conversation. Although the "conference" structure is specifically designed to support group communication and decision making to replace or augment face-to-face meetings, electronic mail systems or document and file transfer-systems can be used in the same way, with the group members rather than the computer sorting and ordering the communications for a single problem or subject. In addition, many special structures or features can be created when the computer is in the communications loop. For example, within a conference structure, a human leader or moderator can be given a very strong role. If there are data as well as qualitative communications involved, ranging from simple yes-no votes to large tables or files of information bearing on a decision, the computer can serve as a decision support tool by analyzing, formatting, and feeding back the

data to the group.

Peter and Trudy Johnson-Lenz (1981:1)) have written about means of structuring computer-mediated communication as "groupware." They assert that:

> For a group to use a computerized conferencing system effectively, it must have some explicit, intentional procedures to follow. These procedures set out the purpose of the group and its tasks, who can communicate with whom and when, how decisions are made and disagreements resolved, the sequence of activities to be used in accomplishing the task, and so forth. The procedures may be norms or rules enforced by the group, or they may include software enforcement. Such procedures constitute a communications structure, without which the group's work will be neither effective nor efficient.

There are thus two main varieties of "structure." Group interaction processes and procedures may be ordered by agreement on norms and roles. The computer may be used to help generate or support such norms or roles, but they depend upon the group members for acceptance and enactment. Secondly, software support may be used to play an active part in the communication. The computer can regulate the flow of communications by, for instance, disallowing private messages among group members, so that all communications are visible to the entire group; enforcing the use of pen names or anonymity; or analyzing and displaying data or responses to surveys or votes.

Background: The Prior Experiment

The first experiment in this series compared the process and outcome of face-to-face versus computerized conferences for two types of

tasks (Hiltz, Johnson, Aronovitch and Turoff, 1980). One type of task was rank ordering or priority setting; this was of particular interest since the problem used has a correct or criterion solution which permits measurement of the quality of decision reached (see below). The form of computerized conferencing used in that study was completely free and unstructured. Surprisingly, we found that although there are significant differences in group process between face-to-face and computerized conferences, the quality of the decision reached was equally good for both media. However, the face-to-face groups achieved higher levels of consensus and greater subjective satisfaction. The greater probability of consensus seemed to be associated with the tendency for dominant persons-- informal leaders-- to emerge in the face-to-face discussions but not in the computerized conferences.

We also noted that the computer conferencing groups appeared to spend a good deal of time trying to communicate about similarities and differences in their rankings for the complex (15-item) ranking task, in order to keep track of where they were in terms of reaching consensus. Lacking the ability to show their lists to one another and to point to items on the list that a group might be developing in common, they seemed to have difficulty deciding how to most fruitfully focus their efforts.

Based on these results, our speculations about the effectiveness of group communication via computer centered on the question of how this communication medium might be improved in terms of the ability to

reach consensus, the quality of the decision and subjective satisfaction. (See Hiltz and Turoff, 1978 for an early discussion of the importance of structuring for group communication effectiveness). Might it help if a process were provided for generating a leader? And could the power of the computer be used to generate displays of data with formatting and analysis that would allow the group to easily view the extent of agreement and disagreement on each of the items being discussed?

The complex ranking task used in the first experiment was "Lost in the Arctic" (see Eady and Lafferty, 1969). It requires the sharing of knowledge by the group members about the usefulness of different kinds of equipment for survival in the subarctic, and their agreement on a rank ordering of the relative importance of the 15 items. This task has a correct or criterion answer produced by the Royal Canadian Mounted Police trained for arctic survival and rescue, and it was selected for use in the second experiment to provide indirect comparisons. Unfortunately, the agreement with the copyright holder specifies that we may use the problem, but not publicly disseminate it.

Both experiments include as dependent variables the ability to reach consensus, the quality of the decision, equality of participation, and subjective satisfaction.

EXPERIMENTAL DESIGN

The objective of this study is to explore the question of whether it is possible to create "groupware" structures to support group decision making that can significantly improve the level of decision quality, consensus, and subjective satisfaction. The mechanisms for structuring the group process chosen are a formal human leadership role and a decision-aid tool based on computer feedback of summarized data on the decision preferences of the individual group members.

A second objective is to increase our confidence in the applicability of our experimental results to managers and professionals. Whereas the first experiment used college students as subjects, in a laboratory setting, this study is a field experiment, with staff members in a variety of organizations serving as subjects in what was termed a "participatory seminar" on computerized conferencing. Organizations included are Banker's Trust, Texas Instruments, and Chemical Abstracts, Inc., among others. Two other changes in procedure were made on the basis of experiences during Experiment 1. The training period was increased from about a half hour to approximately one hour, and included two practice problems as well as free discussion. The maximum time allowed to reach agreement on the Arctic problem was extended from 90 minutes to two hours.

The Independent Variables: Structuring the Group Process

All the groups in this study discussed the "Lost in the Arctic" problem in a synchronous computer conference, in which private messages were not allowed (all items were automatically entered in the group conference) and in which all items were automatically signed with the "real" name of the contributor (no pen name or anonymous entries were permitted). In addition to text communications in "conference comments," a one-line instantaneous or "interrupt" message generated by the computer informed conferees whenever a member changed his or her rank orderings of the 15 arctic items.

A simple four command interface was used (see Appendix). The command "+enter" entered text comments. The command "+order" initiated re-ranking. A list of how far each participant had progressed in the discussion was generated by "+status." Finally, the command "+xpt" put the participant back exactly where she/he had been, if by any chance the connection was lost or the subject otherwise managed to circumvent our software safeguards to keep them within the conference.

Two factors were chosen to vary the structure of the interaction. The first is the selection of a formal group leader. Groups in this "human leader" (HL) condition were asked, after their training

session, to rank order their members in terms of their ability to lead a group discussion. A weighted scoring was used to calculate the results of the group's votes. After their initial ranking for the arctic problem, groups in this condition were told whom they had chosen as leader. The leader's responsibilities were to focus the discussion, suggest specific ranking changes to reach consensus, and summarize the progress. The control groups had a comparable task, rank ordering five candidates for President of the United States, but there was no reporting of the group's choices.

Earlier experimental work on small groups supports our hypothesis that having a leader can increase effectiveness. For example, French (1941) found that groups with leaders were less likely to split into subgroups or factions. Borgatta and Bales (1953) found that a leader was necessary to direct activity and achieve task-oriented goals. Maier and Solem (1952) found that a discussion leader could improve the quality of decision by making sure that potentially valuable minority opinions are taken into account. Palazzolo (1981:217) summarizes by saying that "These and other similar studies indicate that the simple differentiation of membership along leader-follower lines is sufficient and necessary to activate the group membership in the direction of effective goal- and task-directed behavior."

The second factor is the use of the computer to compile, analyze, and feed back to the groups information on the distribution on the rank orderings at different points in time. All groups received a simple text table listing the members' rankings (see Table 1-1). An updated

table was printed before the discussion began, and then every ten to twenty minutes during the two hours of discussion and reranking. The algorithm was that the table printed every twenty minutes by default if it had not been generated sooner. After a table was printed, whenever a member changed his or her ranking, all rerankings during the next ten minutes were collected and a new updated table was then printed. The ten minute interval was arrived at during pretests. New tables every five minutes proved disruptive to the flow of communication. Intervals longer than ten minutes when changes were being made created difficulty for participants in keeping track of the current information. At the time the experiments were conducted, EIES did not have the capacity to support the ability of any conferee to ask for a table at any time, without encountering an unacceptable delay, although that might be a preferable delivery option.

In the "computer feedback" condition, a second table was generated (see Table 1-2). This listed the items in order of their mean ranking by all group members, showed the amount of agreement on each item, and reported two measures of the amount of agreement so that the group could follow its progress toward consensus.

The second feedback table thus provides summarized data rather than the raw data contained in the first. There is supporting experimental evidence in the area of Management Information Systems that summarized data leads to better decisions on the part of individuals than does raw data (Dickson, Senn and Chervany, 1977). In those experiments it was also found that those using the raw data

had more confidence in their decisions. Since our groups had both the raw and summarized data, that potential disadvantage does not appear to be relevant. However, this earlier result was strictly limited to individuals working against a computer model.

From the standpoint of group processes, support for the use of statistical or summarized feedback of opinion oriented data as a mechanism to aid decisions lies in the area of the Delphi Method (Linstone and Turoff, 1975). The classic experiments at RAND (Dalkey, 1969, 1970) strongly support the hypothesis that statistical and controlled feedback of group opinion increases the accuracy of group results. However, these early results were limited to "almanac" type questions (e.g. How long is the Nile river?). The task used in our experiment was much more demanding from the point of view of the group objective and the complexity of carrying out the task. In the Delphi process, the monitor team acted as a leader by filtering out all extraneous information and feeding back only what was determined to be pertinent comments (e.g. "I think Egypt is about 1500 miles long"). Other than the lack of anonymity, our computer feedback condition had many of the characteristics of a real time Delphi (Turoff, 1974).

We thus have a two-by-two factorial design (see Table 1-3). There are six groups per condition, with five members per group. One of the conditions (No Leader, No Feedback) is comparable to the unstructured conferencing condition used in the first experiment.

Table 1-1

Text-Only Table Received in All Conditions

(Example for the Practice Problem)

	ROGER	DOROTHA	DAVID	ANN	CAROLYN
1	C PIE	C PIE	C PIE	E STRAWBERRY	D CAKE
2	D CAKE	D CAKE	E STRAWBERRY	C PIE	B MOUSSE
3	E STRAWBERRY	B MOUSSE	B MOUSSE	D CAKE	C PIE
4	A CREPES	A CREPES	D CAKE	B MOUSSE	E STRAWBERRY
5	B MOUSSE	E STRAWBERRY	A CREPES	A CREPES	A CREPES

Table 1-2

Sample of Computer Feedback Table

(for the Practice Problem)

The overall group agreement is 56.8%.

.-Average Rank for This Item	.-Group Agreement	Item	User's Rank 901	902	903	904	905
1.2	84%	B Mousse	1	1	1	2	1
2.8	32%	E Strawberry	5	4	2	1	2
3.6	64%	C Pie	3	5	3	4	3
3.6	56%	D Cake	2	3	4	5	4
3.8	48%	A Crepes	4	2	5	3	5

Kendall's agreement coefficient is 0.464.

Table 1-3

Design of the Experiment

2 X 2 Factorial

	HUMAN LEADERSHIP	
	YES	NO
COMPUTER FEEDBACK	HLF	NLF
NO COMPUTER FEEDBACK	HLNF	NLNF

6 Groups per Condition

5 Subjects per Group

KEY
HLF= Human Leader, Feedback
NLF= No Leader, Feedback
HLNF= Human Leader, No Feedback
NLNF= No Leader, No Feedback

Dependent and Process Variables

There are two essential dimensions to a successful group decision: solution quality and acceptance or consensus. The quality of a group decision may be assessed by comparing it with objective facts or expert opinions, when they are available. If there is no group consensus or acceptance of a decision, there may not be sufficient committment to motivate its successful implementation.

Total consensus is not necessarily a goal that is related to quality of decision. As Nixon (1979:143) puts it in his summary of small group studies, "conformity and deviance can have either potentially functional or dysfunctional consequences." However, consensus on a decision usually makes the group members feel better about each other and about the decision.

Given these considerations, how can we operationalize criteria for the effectiveness of a group decision support structure? We have conceptualized the following as dimensions to be considered, and they serve as dependent variables in the experiment:

1. Quality of Decision: the average group decision is better than the average of the individual decisions before discussion. This can be measured in terms of a "percent improvement" in the quality of the decision.

2. "Collective Intelligence": the group decision is better than the solution of any of its members before discussion. This is a very strong criterion; past research indicates that "although the group is usually better than the average individual, it is seldom better than the best individual" (Hare, 1976:319).

3. Consensus: although complete consensus is not necessary, there should be enough consensus so that the group can recognize a rough "group decision" that its members are willing to "live with," even if it is not the first choice of all the members.

There are two measures of consensus available from our data: one is the extent of recognition of a group consensus; this is the coefficient of agreement for the "group decision" specified by each member after discussion. The second and stronger criterion might be termed "actual agreement;" it is the level of consensus in the "final individual" post-discussion rankings, where the individuals offer what they "really" think is the best solution, as compared with the solution arrived at by the group.

4. Subjective Satisfaction: How satisfied are the participants with the medium itself, their own performance, and the group interaction?

5. Intervening or Process Variables: In addition to the dependent variables of quality, consensus, and subjective satisfaction, we are interested in several variables having to do with the process whereby these outcomes are reached. For this experiment, we will include the

amount of text discussion, the amount of ranking and reranking activity, and the degree of equality or inequality among the members participating in the discussion. Inequality will be said to occur when one person dominates the discussion. The criterion for "dominance" will be 33% or more of total lines or comments, as compared to the 20% which would constitute an equal share.

A separate content analysis is being performed on the transcripts by Andrew Finn (1982) as part of a Ph.D. dissertation. It will categorize the types of communication which occur and their consequences.

Hypotheses

It was hypothesized that:

1) Human leadership will improve amount of consensus.

2) Human leadership will improve quality of decision.

3) Computer feedback will improve amount of consensus.

4) Computer feedback will improve quality of decision.

5) There will be interaction between human leadership and computer feedback.

6) Human leadership and computer feedback will affect the process of communication as follows:

   a) There will be more re-ranking with computer feedback.

   b) There will be more discussion with human leadership.

7) There will be more inequality of participation with human leadership.

8) Human leadership will be associated with greater subjective satisfaction than computer feedback.

Covariates

Skills and characteristics of the individual will interact with the structure provided and affect the outcome. To the extent that this is true, these factors should be treated as covariates in the primary analyses. For example, if typing speed increases ability to reach consensus, then any observed relationship between Computer Feedback and degree of consensus should be controlled by typing speed, to assure that it is not a spurious relationship caused by differences in average typing ability that were confounded with treatment condition.

The hypotheses below were based on findings and qualitative observations from previous research:

9) Typing speed will be positively associated with quality of decision and ability to reach consensus. (Those with inadequate typing skills will simply not be able to communicate enough in the limited time available).

10) Previous computer experience will be positively associated with quality of decision and ability to reach consensus.

11) Age will be negatively related to quality of decision, ability to reach consensus, lines entered, and subjective satisfaction.

12) Females will be more satisfied with the medium than males.

Another important covariate is "group differences." We did not have anything approaching random assignment to groups for this field experiment, since we were using "real" organizations. Groups at such organizations as Banker's Trust, Kaiser Permanente, and Chemical Abstracts differed not only in terms of the average level of skills related to previous use of computer terminals, but also in the extent to which they were permanent working groups or just a collection of employees of the same organization who did not work together regularly. Therefore, we must also pay attention to "group" as a covariate for our analyses.

SUBJECTS AND PROCEDURE

The participants in this study belonged to organizations which requested a one-day "participatory seminar" (see the announcement and full text of all experimental instructions in the Appendix.) The host organization paid travel and Telenet charges and selected the participants. Following an approximately half-hour face-to-face orientation in the morning, they spent one to one and a half hours learning and practicing on EIES. This practice included a complete single ranking problem. The group ate lunch together and received a "Crib Sheet" of their four commands.

All participants were alone in separate office spaces in the afternoon problem-solving session which followed. After reading the arctic problem and entering their initial rankings they had up to two

hours to reach consensus. Some groups finished five to ten minutes early. A two-hour seminar followed which reviewed the full range of the technology and some applications and impacts. The participating organizations wished to consider office automation applications of computerized conferencing and used the participatory seminar as a means of making a more informed decision. In the list of participating organizations below, those with asterisks did subsequently decide to take some memberships in EIES and try an application. Thus, the participants did not define themselves as subjects in an experiment, but rather as a group of colleagues trying out a technology which they might decide to use more permanently. The following list of runs used for this report also shows the number of groups from each organization included in the experimental data and geographic locations.

Kaiser Permanente * (2), Portland, Oregon
Foundation for the Arts (New York based organization; experiment conducted at Upsala College in East Orange, NJ)
George Washington University (3), Washington, D.C.
Chemical Abstracts (2), Columbus, Ohio
Banker's Trust Company * (4), New York, New York
Texas Instruments (2), Dallas, Texas
North American Phillips (2), New York, New York
General Accounting Office (2), Washington, D.C.
Stanford University, Stanford, California
State of Florida, Department of Higher Education, Tallahassee, Florida
American International Insurance Groups, New York

CERN (Consumers Education Resource Network) * (2), Rosslyn, Virginia
New Jersey Institute of Technology * (1), Newark NJ

We are also grateful to the Defense Communication Agency, Reston, Virginia, where we conducted three pretest runs which resulted in our making some final modifications to the experiment. Several runs had to be deleted from the experimental data base either because one of the group members had previously seen the arctic problem, or because the system crashed.

Since we used groups consisting of employees within an actual organization, we were not able to choose subjects for random assignment to groups. A kind of modified systematic random sampling technique was used to assign groups to condition. The conditions HLF and NLNF were paired, as were HLNF and NLF, since each of these involved one condition with feedback and one without, and one condition with human leadership and one without. We chose the initial condition randomly. Then we proceeded to assign groups to conditions according to two principles:

Fill in the condition which now has one less run than the others;

Assign groups from organizations with two groups to one of the "paired" conditions.

The experiment was automated. All subjects proceeded through 57 steps, led by the computer. Methodological details about the use of

EIES in conducting the experiment are described in a subsequent chapter.

Description of the Analysis of Variance Designs

The basic method used to analyze the data is an "analysis of variance." This analysis partitions the total variance of the dependent variable into treatment and error variance. In comparing groups that received different treatments, we are attempting to see if there are significant differences "between groups" associated with different treatments in the experiment. The independent variables are Leadership and Feedback; we also examine whether there is significant interaction between the two. Another factor to consider is group differences. Finally, a "nested" design in which individual observations are nested within their group allows us to see if observed differences among treatments are significant when the effect of variations among the groups is removed.

Two designs for the analysis of variance are used. The individual level of analysis uses the 120 subjects as independent observations and measures the significance of group differences. Unfortunately, there are significant differences associated with "group" for almost all variables. The group level uses a nested design, and uses group averages and performance parameters rather than individuals. The level of significance adopted is .05, but differences with less than a .10 level of significance will also be reported.

SUMMARY

In this second of a series of three controlled experiments on group decision making via computerized conferences, Leadership and Computer Feedback are used in a two-by-two factorial design to vary the structure of the conferencing medium. Dependent variables include measures of the process and outcome of the groups' conferences on a rank-ordering task.

That this is a field experiment, carried out with "real" groups of managers and professionals, is perhaps its greatest strength and its greatest weakness. Because we used actual groups of employees in existing organizations, who participated in their office settings rather than coming to a laboratory as "subjects," we may feel more confident about generalizing our findings to the "real world" of the office. At the same time, the fact that we used naturally occurring groups and subjects in their everyday settings means that we had less "control" over the experiment. The groups are not similar, and constitute a source of variance that may be stronger than our experimental manipulations in the structure of the conferencing capability. If we had used random assignment to experimentally constituted groups in a laboratory setting, we may have found more statistically significant differences associated with our structural variations.

CHAPTER TWO

QUALITY OF DECISION

MEASURES OF QUALITY OF DECISION

The Decision Data

The ranking problem which the groups were trying to solve is called "Lost in the Arctic." Because of proprietary agreements, we cannot reproduce it in its entirety. The situation is that the group has crashed in a remote subarctic region. They have pulled a pile of 15 items out of the wreckage of the plane before it sank. Their task is to reach agreement on the relative importance to their survival of the 15 items. They may not just decide to "take" or "leave" the items, but must arrive at agreement on a common rank ordering. Thus, though the situation is purely fictional, the problem is an example of the kind of priority setting and planning for resource allocation in which management groups must frequently engage. The subjects were instructed to think of the problem in these terms (as an exercise in reaching agreement on priorities) and there were no complaints about the irrelevancy of the particular ranking problem chosen.

The ranking problem was formally answered three times. Each group member read the problem and individually gave an initial answer. At the end of the two-hour time limit for the group discussion, or when the group announced that it had reached agreement, each person was asked to report the agreed upon common group ranking, or their

impression of the "group decision" at that time. Each person also entered his or her "final individual" decision, the rank ordering which was really considered to be the best answer, having discussed and thought about the problem for two hours. In addition, individuals were free to re-rank at any time. We can recover the stored information on these rankings for any individual at any point in the discussion, compute an average group answer for any point, or count the number of rerankings done.

The criterion is the solution offered by the experts, the Canadian Royal Mounted Police, who are trained and experienced in rescue in the subarctic area in which the fictional plane crash occurred.

Following the procedure established in previous studies using this problem (see Eady and Lafferty, 1969), correctness or quality of decision is computed for each individual for each ranking by subtracting the given rank from the correct rank. For example, one item is snowshoes. If the correct rank for snowshoes were 10 and the person put them in fifth place, this would be a deviation of 5. Signs are ignored (a +5 is the same as a -5) and the sum of the deviations of the 15 items from the correct ranking is the individual's "deviation score." Thus, the smaller the "deviation score," the better the solution.

The Percentage Improvement Measure

Groups and individuals varied in terms of their prior knowledge about

the items in the problem, and some began with much better solutions, before discussion, than others (See Table 2-1). There are significant differences among groups in the quality of the initial, pre-discussion rankings. These group differences are not associated with experimental condition. Thus, in all analyses, we must account for initial differences mathematically so that we can compare relative improvement due to discussion, not just the absolute quality of the decisions. The method which we have adopted to handle this problem is to compute a percentage improvement, calculated as Initial (pre-discussion) deviation minus Group solution deviation/Initial deviation. This lets us compare relative improvements, regardless of initial differences in quality of solution among the groups. Percentage improvement will be our primary measure of the quality of the groups' decisions.

"Collective Intelligence"

A second, very stringent measure of improvement will be examined briefly at the end of this chapter. We have defined "collective intelligence" as the ability of a group to arrive at a solution that is better than any of them could have achieved individually. This will be determined by comparing the deviation scores of the best group member before discussion with the group decision. If the group's decision is better than that of its "best" member, it will be said to have achieved "collective intelligence."

Table 2-1

Mean Initial Deviation Scores by Condition

Condition	Feedback	No Feedback	All
HL	53.8	51.7	52.7
NL	49.7	54.5	52.1
Both	51.7	52.4	52.4
			(SD= 13.1)

ANOVA, Individual Observations (N=120)

Leadership F=.1, NS

Feedback F=.6, NS

(Leader x Feedback) F=2.1, NS

Group F=1.9, p= .02

ANOVA, Group Level (N=24)

No significant differences

DIFFERENCES IN QUALITY OF THE GROUP DECISION

Table 2-2 examines the data on the absolute quality of the group decision. There are small but consistent, statistically significant differences in quality of group decisions in favor of those groups which had leaders, and against those with Feedback, when the data are examined in terms of 120 individual scores. However, by far the strongest source of variation has to do with differences among groups. When the variance associated with group is used as an error term, and the 24 group scores are used as the basis for analysis, there is no significant difference whatsoever.

The percentage improvement data are shown in Table 2-3. Here, the initial differences in quality of decision before discussion are eliminated. The quality of decision of groups in all conditions tended to improve noticeably. However, there are significant differences associated with feedback, and the interaction among leadership and feedback, when the 120 individual scores are examined; the NLF condition improved much less than any of the others. None of the other differences are significant. Once again, however, the strongest, most significant differences are associated with group; some groups were much better than others, regardless of condition. When the group differences are used as an error term for the second analysis, there are no significant differences among conditions.

What is it that is making some groups much "better" than others,

independent of condition? Clearly, it must have something to do with group composition. Analysis shows that among the group composition variables which appear to explain much of the apparent differences in quality of group performance are the quality of the leader's own decision, if there is a leader; the quality of the "best individual's" pre-discussion score; and attributes such as age, sex, and typing, which are related to the process and outcome of the group's decision making process. We will first examine the attributes of leaders, and how they affected the quality of group decisions.

Table 2-2

Mean Group (Post- Discussion) Deviations from Correct Answer, by Condition

Condition	Feedback	No Feedback	All
HL	35.4	34.1	34.7
NL	38.5	35.7	37.1
All	37.0	34.9	35.9
			(SD= 2.7)

ANOVA, N=120

Leadership F=22.9, p=.001

Feedback F=18.0, p=.001

Leadership X Feedback F= 2.4, p= .12

Group F= 89.1 p=.001

ANOVA, Group Level (N=24)

Leadership F= .26, NS

Feedback F=.20, NS

L X F F=.03, NS

Table 2-3

Percentage Improvement in

Deviation from Criterion, by Condition

(Individual Deviation- Group Deviation/ Individual Deviation)

Condition	Feedback	No Feedback	All
HL	30.8	31.0	30.9
NL	16.0	31.0	24.6
All	23.4	32.1	27.7
			(SD= 20.2)

ANOVA, Nested Design (N=120)

Leadership F= 2.89, p= .09

Feedback F= 5.57, p= .02

Leadership X Feedback F= 5.32,p= .02

Group F= 4.09, p= .001

ANOVA, Group Level (N=24)

Leadership F= .71, NS

Feedback F= 1.36, NS

Leadership X Feedback F= 1.30, NS

The Selection and Performance of Leaders

Each participant in a leadership condition ranked the group members from "1" (highest) to "5" (lowest) in terms of their ability to lead a group discussion in this medium, following the practice session. The correlation between number of lines entered during the practice discussion and leadership ranking was -.44 (p=.001); that between number of comments entered during the practice and leadership ranking was -.46 (p=.011). Those who entered the most during the practice were those who were ranked highly (1 or 2) and selected as leaders.

The deviation from the correct decision varied greatly among leaders, from a low of 30 to a high of 76. There was absolutely no correlation (Pearson's of .01) between the quality of the leader's initial pre-discussion solution to the problem and the likelihood of being selected as a leader. Thus, we see that it is the relatively verbose person who became leader, not the person with the most knowledge about the problem the group would try to solve.

Those who were ranked high in the leadership selection continued to be the most active participants in the discussion on the problem. The correlations are as follows:

Number of run lines entered: -.50, p=.001

Number of run comments entered: -.41, p=.001

Percent of all lines entered: -.59, p=.001

% of total comments entered during the run (problem discussion): -.51, p=.001

Clearly, leaders were having a disproportionate influence upon group decisions. But some of these leaders had "correct" opinions about the solution to the problem, and some were incorrect. Looking at the group level data, there is a high correlation between the quality of the leader's pre-discussion decision, and the absolute quality of the group decision reached (Pearson's R= .71, p=.001). In terms of percentage improvement, the correlation is .54 (p=.001). Thus, we see that for those groups which did have a leader, much of the variance in the quality of the group decision is explained by the chance of whether or not they happened to choose a leader who was knowledgeable about the problem and who would influence the group to make a good decision rather than a poor one.

Influence of the "Best" Member

Groups also varied greatly in terms of the presence or absence of one or more persons who started out fairly knowledgeable about the problem; the range of initial deviation scores for the group member with the best pre-discussion solution ranged from 24 to 54.

The chance of having one or more members with an initially good solution was not distributed evenly among the groups. Having such a member did influence the quality of the final group decision: the correlation between the deviation of the group's best member ("Least

Deviation") and the quality of the group decision is .44 (p= .001). However, unlike the leader's initial opinion, there is no correlation between the quality of the best member's pre-discussion solution and the percentage improvement in the group; it is apparent that the opinion of the "best" member of the group did affect the absolute quality of the group decision, but it did not have a disproportionate impact on that decision, as did the opinion of an elected leader.

If "Least Deviation" is used as a covariate, with either absolute quality of decision or percentage improvement as the dependent variable, then there are no significant differences among groups associated with condition. However, those groups in the Feedback conditions still appear has having noticeably smaller percentage improvements when Least Deviation is covaried out, though not statistically significant (p= .20). Thus, there is once again the suggestion that feedback is detrimental to reaching high quality decisions, but this effect is dependent upon how quality of decision is measured and what other variables are taken into account.

The Effect of Group Composition

Several demographic and skill characteristics are related to quality of group performance, and are unequally distributed among groups. This would be expected in any collection of "real" staff groups.

Typing skills are significantly related to many aspects of participation level and effectiveness (see Chapter 4 on group process

variations for details). But typing skills are not evenly distributed across conditions: there is a higher level of typing skill among those in the Human Leader conditions.

Age is another variable that generally correlates with level of performance and subjective satisfaction-- in this case negatively, with the older subjects doing more poorly. (The correlation between age and individual percentage improvement is -.19, p=Ø4). And age is higher for the two feedback conditions.

Sex composition is also strongly related to improvement. With males coded as "1" and females as "2," the point biserial correlation between sex and individual percentage improvement is .21 (p=.Ø2), and there are significantly more females in the No Feedback conditions.

Previous computer experience, related to better performance, is significantly higher for the HLNF condition than for the others. Education levels are higher for the No Feedback conditions.

Thus, we see that all of the demographic characteristics associated with more improvement in quality of decision are skewed in favor of the groups in the human leader conditions and/or against those with the Feedback condition. These correlations support the observation reported above that "something" related to differences among the groups themselves, rather than the experimental condition, explains much of the apparent poor quality of decisions observed for the NLF groups.

"COLLECTIVE INTELLIGENCE," BY CONDITION

Table 2-4 shows the extent to which the groups were able to incorporate and surpass the knowledge of their "best" member in making a collective decision, by condition. Overall, half of the groups succeeded in reaching a decision that was better than that which could have been made by any individual member without benefitting from the knowledge and insights of the other members. This is an encouraging result for the effectiveness of computerized conferencing as a means of communication, since previous studies have found that such "collective intelligence" rarely occurs (Hare, 1976).

Looked at purely in terms of a "yes - no" dichotomy, it appears that feedback is detrimental to the emergence of "collective wisdom." In the Human Leader, Feedback condition, two of the six groups produced a group decision better than that of their best member. For No Leader, Feedback, only one out of six accomplished this. Turning to the No Feedback conditions, four out of six with a leader surpassed their best member, and one reached the level of the best member; with No Leader and No Feedback, five out of six were better, and one group decision was equal in quality to that of the best member. It appears that the feedback tables are having the effect of decreasing the influence of the most knowledgeable member, perhaps by creating pressure to reach a compromise rather than exploring the reasons underlying a "deviant" member's opinion, which may in fact be

superior.

This examination of the differences in the frequency of achieving "collective wisdom" was followed by an analysis of variance which uses as a dependent variable how MUCH better or worse the group decision is than the opinion of the best member. The dependent variable is the deviation score of the group decision from criterion, minus the deviation of the best member's pre-discussion opinion. Thus, a negative score indicates a group decision that is better (closer to the correct answer), and a positive score indicates that the group decision is worse, in that it deviates more from the correct decision. This analysis confirms that the groups with feedback are significantly less likely to achieve "collective intelligence."

Table 2-4

"COLLECTIVE INTELLIGENCE," BY CONDITION

LEAST DEVIATION OF ANY INDIVIDUAL (LD) VS. GROUP DEVIATION (GD)

	HUMAN LEADER			NO HUMAN LEADER	
	LD	GD	**	LD	GD
	--	--	**	--	--
			**		
FEEDBACK	34	38.4	**	26	29.2
	42	52.8	**	38	70.0
	*30	22.0	**	*30	28.8
	38	37.6	**	24	30.4
	32	37.2	**	32	38.0
	*38	24.4	**	38	34.0
			**		
**********					
**********					
			**		
NO FEEDBACK	*32	24.0	**	*28	24.8
	*40	32.0	**	*54	35.2
	46	52.0	**	*38	34.0
	*34	32.0	**	=42	42.0
	=26	26.0	**	*50	36.0
	*48	38.4	**	*52	44.0

* indicates LD < GD
= indicates LD = GD

ANALYSIS OF VARIANCE
(Dependent Variable= Group Dev- Dev Least)
(Means by Condition)

	Leader	No Leader	All
Feedback	-.27	7.02	3.47
No Feedback	-3.60	-8.00	-5.80
All	-1.93	-0.40	-1.00
			(SD=10.6)

Leadership, F= .17, Not Sig
Feedback, F= 6.20, p= .02
L x F, F= 2.54, p= .13

SUMMARY

In examining the percentage improvement of the average "group" decision compared to the average for the five members before discussion, we find that the average improvement is about 28%. When group differences are not taken into account, human leadership appears to improve decision quality and computer feedback appears to decrease decision quality. However, these apparent differences tend to disappear when differences among the groups are controlled. Differences in group composition are a more powerful determinant of differences in percentage improvement than are the experimentally induced differences in the structure of the communication medium.

Groups selected leaders on the basis of their performance during the practice session, rather than on the basis of their knowledge about this particular task. Leaders tended to be those who were the most active participants in the practice session, and continued to contribute a disproportionate number of comments to the discussion during the problem session. Some leaders happened to be knowledgeable about the problem, and others were not. There is a very high correlation (Pearson's R= .74) between the quality of a leader's pre-discussion solution and the quality of the group decision reached.

Groups also varied markedly in the extent to which they started out with one or more knowledgeable members, and to which they were

composed of members with characteristics related to the emergence of better decisions. For instance, groups with more women did better. When these large differences in group composition are taken into account, there is no significant difference among conditions in percentage improvement in quality of decision.

Turning to the "strong" criterion of "collective intelligence" (a group decision which is better than the decision which would be made by the most knowledgeable member acting individually), there are statistically significant differences among conditions. Those groups with computer feedback were less likely to attain collective intelligence.

Taking into account the various measures of quality of decision and covariates examined, one reaches the overall conclusion that the primary determinants of the quality of the group decision will be the quality of the best member's pre-discussion solution; the quality of the leader's solution, if there is a leader; and attributes such as sex and previous computer experience of the group members. However, there is also a fairly consistent tendency for the presence of computer feedback to be detrimental to a high quality group decision. The feedback tables appear to decrease the influence of the "best" member, by creating pressure to compromise.

CHAPTER THREE

ABILITY TO REACH CONSENSUS

Consensus was measured by using Kendall's coefficient of concordance for the five rankings reported by each individual in each group. Kendall's varies from Ø for no agreement to 1.ØØ for perfect agreement on the placing of the 15 items ranked by the group. We computed Kendall's for four points in time:

INITIAL= the initial, pre-discussion rankings
DISCUSSION= The rankings which existed in the last table generated before the end of the discussion.
GROUP= the rank orders produced after discussion which was their "perception of what the group decided."
INDIVIDUAL= the final post-discussion according to what "you, yourself, really think the proper ranking of the items should be, now that you have had the discussion."

There is no significant difference in the initial levels of agreement before discussion (Table 3-1). The average coefficient of .55 before discussion shows that the groups did have considerable "work" to do in order to reach agreement.

REACHING A GROUP DECISION

At least 95% agreement was reached, on the average, in all conditions. The levels of agreement in all conditions are so high that the differences which do occur are not statistically significant. As shown in Table 3-2, however, there are some

interesting qualitative differences. In the condition with a human leader and no feedback, five of the six groups reached 100% agreement. The condition with both a human leader and feedback was the worst; none of these groups reached 100% agreement. These qualitative differences are reflected in the fact that the effect of the interaction between leadership and feedback is "almost" significant, at .08. The lack of significant association between condition and Group Kendall's did not change when the Initial Kendall's was used as a covariate.

These results vary from those of the first experiment, where computerized conferencing groups did not reach such high levels of agreement on a group decision, and none were able to reach 100% agreement on the arctic problem. The differences may be attributable to any of five factors:

1) The groups were allowed two hours, rather than only 90 minutes.

2) All groups received a practice ranking problem. They also had a longer training time (over an hour, as compared to less than half an hour in the first experiment).

These changes were made because it was observed during the first experiment that groups seemed rushed by the 90 minute deadline, and that some individuals needed more learning time than had been provided. We also know from previous experiments that training does help group performance, so that it can be expected that having practiced two rank ordering tasks, the subjects would be more comfortable and familiar with the procedure.

3) All groups did have the text-only tables and notification of ranking changes as they were made, so that they did not have to separately change their ranks and communicate these changes to one another. In the first experiment, tables of ranks were made available only at the beginning of the discussion.

4) These were more nearly "real" groups; they were familiar with one another as members of the same organization. Thus, it can be expected that they would find it easier to work together and reach agreement than did the groups of strangers used in the first experiment.

5) The subjects had more previous computer experience than did the subjects of the first experiment; as we will see below, this factor is related to ability to reach consensus.

Table 3-1

Initial (pre-discussion) Agreement, by Condition

(Kendall's Coefficients of Consensus; 1.00= 100% Consensus)

ANALYSIS OF VARIANCE (N= 24 GROUPS)

(two-by-two factorial)

	F	NF	All
HL	.51	.56	.54
NL	.57	.57	.57
All	.54	.56	.55

Human leadership: F= .44, p= .51 (NS)

Feedback: F= .18, p=.68 (NS)

Leadership X Feedback: F= .38, p=.55 (NS)

Table 3-2

Degree of Consensus on Group Decision

(Kendall's Coefficients of Consensus; 1.00= 100% Consensus)

ANALYSIS OF VARIANCE (N=24 GROUPS)

(two-by-two factorial)

	F	NF	All
HL	.953	.997	.975
NL	.986	.960	.972
All	.969	.978	.973

Human leadership: F=.02, p=.90 (NS)

Feedback: F= .22, p=.64 (NS)

Leadership X Feedback: F=3.36, p=.08 (NS)

GROUP SCORES BY CONDITION

	HL	NL
Feedback	.998	1.000
	.995	1.000
	.994	.988
	.972	.982
	.900	.974
	.867	.969
NF	1.000	1.000
	1.000	1.000
	1.000	1.000
	1.000	.988
	1.000	.956
	.982	.813

RESULTS FOR INDIVIDUAL RANKINGS

The final group rankings represent the ability of the group to arrive at a nominal consensus, perhaps involving compromises among underlying disagreements. The amount of agreement among the final individual rankings, which represent the "real" opinions of the individuals, may be a better measure of actual agreement or consensus.

As shown in Table 3-3, there was also no significant impact of human leadership or computer feedback on the ability of individuals to reach a genuine consensus after a computerized conference. The consensus among individuals is lowest, on the average, for groups with neither Human Leadership nor Feedback, but by only about five points on the Kendall's scale.

The last table in this chapter (3-4) shows the analysis of variance for the last rankings by the subjects during the discussion. Here, we do obtain some statistically significant differences. Either human leadership alone, or computer feedback tables alone, aided consensus. However, in combination they canceled each other out and were no better than a structure without either aid.

Table 3-3

Degree of Consensus on Final Individual Ranking

(Kendall's Coefficients of Consensus; 1.00= 100% Consensus)

ANALYSIS OF VARIANCE (N= 24 GROUPS)

(two-by-two factorial)

	F	NF	All
HL	.864	.879	.871
NL	.859	.829	.842
All	.861	.854	.858

Human leadership: F=.34 (NS)
Feedback: F= .03(NS)
Leadership X Feedback: F=.24 (NS)

****

Table 3-4
Degree of Consensus Among
Last Subject Rankings During Discussion
(Kendall's Coefficients of Consensus)

	F	NF	All
HL	.854	.980	.917
NL	.929	.849	.889
All	.891	.915	.903

Leadership, F= .5 (NS)
Feedback, F= .35 (NS)
Leadership x Feedback, F= 6.92, p= .02

FACTORS RELATED TO THE ABILITY TO REACH CONSENSUS

As would be expected, there was some relationship between degree of initial agreement and degree of agreement on the final group decision (r= .46, p= .Ø3). However, using Initial Kendall's as a covariate did not change any of the relationships examined.

Those groups with the highest levels of agreement also tended to have better decisions. The Pearson's correlation coefficient between the final Kendall's for group decision and final deviation (from criterion) scores is only -.18, however, and not statistically significant. On the other hand, the correlation between the final individual Kendall's and the quality of the final individual rankings is much stronger (r= .55, p= .Ø1). The difference in these two relationships suggests that "real" agreement is positively related to good decisions, but that compromise in "real" opinions in order to reach group consensus also compromises quality somewhat.

There was no relationship between degree of initial pre-discussion agreement and final quality of group decision (r= -.Ø4).

All items on the post-experimental questionnnaire were correlated with the Kendall's coefficient measure of consensus for the initial, final group, and final individual rankings. The perception of the degree to which the medium is satisfactory for giving and receiving orders is significantly related to the group consensus score (r= .42,

p= .04), as are several other items in the set of questions on perceptions of the medium. There was also a strong relationship with perception that the group had reached consensus, (r= .68, p= .001), which does serve as a measure of consistency between the subjectively reported and objectively measured performance of the groups. Those groups for which the final individual rankings were most similar felt most "productive" (r= .40, p= .05).

Looking at group composition, ability to reach consensus was negatively related to age (Pearson's r= -.24 for group consensus, not significant; -.42 for final individual consensus, p= .04). There was a strong correlation between typing ability and previous experience with computers and the ability of the group to reach consensus. (The Pearson's correlations are .52 for typing ability and .62 for previous experience with computer terminals, both statistically significant at the .01 level).

Typing was measured on a four-point scale: hunt and peck, rough or casual typing, good typing (30 wpm, error free), and excellent typing. Past use of computer terminals, for any kind of application, was self-reported as never, once or twice, three-ten times, or frequently. Since the measure of correlation used is Pearson's, the square of the coefficient is the proportion of variance in the dependent variable explained by the independent variable. So, for instance, our correlation of .62 between previous experience with computer terminals and the amount of agreement among the final group rankings can be interpreted to mean that almost two-fifths of the

variance in the amount of consensus can be predicted on the basis of previous use of computer terminals. In other words, previous individual experience is strongly related to the effectiveness of computerized conferencing to reach consensus. Typing ability and previous experience with computer terminals are interrelated, as would be expected (r=.39).

SUMMARY

Groups in all conditions were able to reach high levels of group consensus. There were no consistently significant differences among conditions. The Human Leadership, No Feedback condition is best for obtaining 100% agreement. In terms of agreement reached during the discussion itself, either the human leader, alone, or computer feedback, alone, were effective. The worst condition appears to be the combination of human leader and computer feedback.

Compared to the weak and inconsistent relationships found for the structural variations, social-psychological attributes of the groups and individuals are stronger predictors of ability to reach consensus. Strong correlations between group consensus scores and both typing ability and previous use of computer terminals demonstrate that the medium is more effective for novice groups whose members have some related skills and previous experience.

The main contrast is not among conditions, but with the outcomes for the same problem obtained in the first experiment in this series. In

those computer conferences, the participants were unable to reach very high levels of agreement on the arctic problem. The differences which may be important are adequate practice time, adequate time to complete the task when using a new medium, previous experience working with computer terminals, and a pre-existing identity as members of the same organization.

CHAPTER FOUR

VARIATIONS IN GROUP PROCESS

In this chapter, we will look at how Computer Feedback and Human Leadership affected the process of the group discussion. The process variables measured are the amount of text discussion, the amount of ranking, and equality of participation in the text discussion. In addition, we will examine the extent to which subject characteristics-- age, sex, education, typing ability, and previous computer experience-- affected performance or process variables. Finally, we will see if there is a relationship among the the process variables and the outcome variables (consensus and quality of decision).

DOMINANCE AND STRUCTURE

Dominance was defined as one person contributing much more (33% or more) than an equal share of the discussion, whether measured in terms of percentage of total lines or percentage of total comments.

Dominance rarely occurs in synchronous computerized conferences, regardless of condition. Only four groups had a dominant person measured in terms of percentage of lines, and one of these occurred in each condition. Only three individuals contributed over 33% of the comments for their conference. Thus, there is no relationship between structure and dominance for this experiment. It could be

that structural effects on dominance would be observed if we had a larger group and a long-term asynchronous conference, or if we had implemented our structural variations differently.

THE EFFECTS OF HUMAN LEADERSHIP AND FEEDBACK ON AMOUNT OF DISCUSSION AND RANKING

It was hypothesized that the presence of the human leader would result in more talk and less reranking; whereas the presence of the computer feedback tables would result in more rerankings at the expense of text discussion. The fact that the extra table of feedback data is printing every ten minutes may cut down on probable discussion. It is if the computer becomes a participant in the group, with its entries followed by a spate of reactions, in the form of changes in the numbers summarized in the tables, rather than in the form of text comments to other members.

The data indicate support for the hypothesis that the feedback tables decrease amount of discussion in terms of number of comments (Table 4-1). The feedback tables were present for the practice problem, and for both the practice problem and the arctic problem, they were significantly associated with fewer comments per (human) participant. However, there is no support for the idea that the human leader would encourage more discussion.

In Table 4-2 we see that there were significantly fewer rerankings when there was a human leader. (This table, and the one at the bottom of 4-1, are shown at the group level of analysis because "group" was significantly associated with both variables when analyzed at the individual level). It appears that the leader tries to have

discussion and agreement on rerankings that group members will do to reach consensus; when there is no leader, the individuals are more likely to independently do rerankings whenever they change their minds. Contrary to our hypothesis, there was not significantly more re-ranking associated with feedback. There is an indication that feedback makes no difference when a leader is present; and that when there is feedback but no human leader, the most reranking occurs, but this is not statistically significant.

Table 4-1

VARIATIONS IN NUMBER OF COMMENTS, BY CONDITION

Mean Number of Practice Comments, by Condition

Analysis of Variance

	Feedback	No Feedback	Both
Leader	7.1	9.9	8.5
No Leader	7.8	8.3	8.0
Both	7.5	9.1	8.3

ANOVA, Individual Level (N=120)
Leadership, F=.74, NS
Feedback, F=9.06, p=.01
Leadership*Feedback, F= 4.11, p=.05
Group, F= 1.47, p= .11

Mean Number of Run Comments, by Condition

	Feedback	No Feedback	Both
Leader	16.2	20.2	18.1
No Leader	16.5	21.5	19.0
Both	16.4	20.8	18.6

ANOVA, Group Level (N=24)
Leadership, F=.23, NS
Feedback, F= 6.71, p= .02
Leadership*Feedback, F= .07, NS

Table 4-2
Mean Number of Run Re-Rankings, by Condition

	Feedback	No Feedback	Both
Leader	3.4	3.5	3.5
No Leader	5.4	4.8	5.1
Both	4.4	4.2	4.3

ANOVA, Group Level
Leadership, F= 11.38, p= .Ø1
Feedback, F= .3, NS
Leadership*Feedback, F= .68, NS

THE EFFECTS OF SUBJECT CHARACTERISTICS ON PERFORMANCE

In Table 4-3, we see that several characteristics of the subjects were related to their apparent facility with use of the system. The older participants started more slowly, writing fewer lines during the practice. Their percentage of the total lines written during the problem discussion ("run") was also smaller than that of younger subjects, but not quite at the .Ø5 level of significance. Their typing ability was also poorer, and their solutions improved poorer as a result of the discussion.

Women wrote more comments than men. This is probably confounded by the fact that the women were better typists. Their solutions also improved more than those of the men.

Those with higher levels of education wrote more lines and comments.

There were no other correlations with general educational level.

Typing ability was positively related to the number of lines written during both the practice and the run. Since it was not significantly related to the number of comments, this means that those with poorer typing ability tended to make the same number of comments, but to keep them much shorter in order to minimize typing.

Previous computer experience was strongly related to many aspects of performance, including the number of lines written, and the proportion of all lines and comments written. However, it was not related to improvement in the quality of the decision as a result of the discussion.

Table 4-3
Significant Pearson Correlation Coefficients (p= <.05)
Between Subject Characteristics and Performance Variables

VARIABLE	AGE	SEX	ED	TYPING	COMP
PLINES				.32	.20
PCOMMENTS					.41
RLINES	-.21		.25	.34	.24
RCOMMENT		.21	.20		.24
PRRANKS					.22
LINESPER				.25	.30
COMMPER					.28
% IMPROVE		.19			
IND IMP	-.19	.22			

KEYS

ED= Educational level

TYPING= typing skill, self-rated

COMP= previous computer terminal experience

PLINES= number of lines of text entered during practice

PCOMMENT= number of comments entered during practice

PRRANKS= number of re-rankings during practice

RLINES= number of lines entered during run (problem solving session)

RCOMMENT= number of comments entered during run

LINESPER= subject's lines as a percentage of total group lines entered during run

COMPER= subject's comments as percentage of total number of comments entered by group

% IMPROVE= Percentage improvement in group solution compared to individual pre-discussion solution

IND IMP= Initial Individual deviation from criterion-final individual deviation/initial deviation

PROCESS VS. OUTCOME

We have observed many statistically significant relationships among condition, subject characteristics, and such process variables as the number and percentage of comments entered in the discussion. However, there is no significant relationship between comment or ranking behavior, and improvement in the quality of decision. There was a weak but significant relationship between number of run lines and ability of the group to reach consensus ($r=.18$, $p=.05$). Thus, though we have been able to demonstrate that the different structures resulted in somewhat different behavior patterns among the subjects, this did not have much significance or importance in terms of the success of the group process, for this experimental task.

Perhaps there are more qualitative differences in group process created by the structures we implemented which are related to quality of decision or consensus. Andrew Finn has undertaken a content analysis of the transcripts of the discussion for his Ph.D. dissertation (Finn, 1982). These results will be disseminated when available. Among the types of content that will be coded are attempts to organize the survival situation, attempts to organize the group's discussion, and "position dependent" approaches which address the "numbers" or "ranks" to be assigned to items as a way of handling the task.

SUMMARY

There is a small but statistically significant tendency for less discussion with feedback tables. With a human leader, there is less re-ranking, but the presence of feedback tables has no significant effect on the amount of re-ranking activity.

Many subject characteristics were significantly related to measures of performance. Older subjects had poorer typing ability and improved their solutions less as a result of the discussion. Those with higher levels of education wrote more comments. Previous computer experience was related to contributing a larger proportion of the discussion. Females, who also had better typing skills, contributed more comments than males.

Though amount of text entered and re-ranking frequency are related to experimental condition, they are not related to differences in improvement in the quality of the decision. Only number of lines entered is related to ability to reach group consensus, and this is a weak relationship.

CHAPTER FIVE

SUBJECTIVE SATISFACTION

The post-experimental questionnaire included questions on a number of different aspects of subjective satisfaction of the participants. In this chapter, we will look at how subjective satisfaction varies according to experimental condition and characteristics of the subjects.

The first set of questions had to do with the problem; generally, the ratings were positive in terms of its being interesting, realistic and clear. This was followed by a series of 7-point semantic differential scales originally designed by the Communications Studies Group in Great Britain for their experiments with group discussions via various communications modes (see, for instance, Short, Williams and Christie, 1976). These questions ask the participants to rate the medium itself, from completely satisfactory (1) to completely unsatisfactory (7) in terms of how satisfactory it is for specific kinds of communication activities. The items and the means are shown below, arranged from those functions for which the participants saw the medium as most satisfactory to those for which it was perceived as least satisfactory.

	Exp2	Exp1
Exchanging opinions	2.7	3.5
Giving or receiving information	2.8	3.6
Problem solving	2.8	4.4
Generating ideas	3.0	3.1
Giving or receiving orders	3.0	3.2
Bargaining	3.8	4.4
Persuasion	4.0	4.1
Resolving disagreements	4.1	4.5
Getting to know someone	4.3	3.9

Except for "getting to know someone," the ratings of the medium by the subjects in this experiment are consistently higher than those for the first experiment. The explanation for the generally higher ratings is probably the longer training time and generally higher levels of previous experience with computer terminals. Ratings for "getting to know someone" may be lower because the subjects in this experiment generally knew one another beforehand, whereas those in the first experiment were generally strangers. One cannot accurately report the extent to which a medium is satisfactory for "getting to know someone" if the other participants are previously known.

The next set of questions dealt with the group discussion itself and the participants' experiences and perceptions of it. They were asked to rate the discussion in terms of how pleasant it was, how satisfied they were with their own performance, whether or not the group

reached a consensus, whether they agreed with the group decision, and whether or not the general feeling of the group was friendly, interested, and productive (see Appendix for complete wording and distribution of responses).

EFFECTS OF EXPERIMENTAL CONDITION ON SUBJECTIVE SATISFACTION

Using analysis of variance at the individual level and cross-tabulations, we found that, generally, the differences among the conditions are not statistically significant. The exceptions are as follows:

- .The issues seemed less clear when there was a human leader. (HL mean= 2.8, NL= 2.3, p=.Ø3)

- .For "giving and receiving information," there was an interaction between Human Leadership and Feedback, significant at the .Ø2 level. Human Leadership with Feedback received the highest rating (mean= 2.4), while HLNF received the poorest (mean= 3.2).

- .For "getting to know someone," the NL conditions were rated more highly than the HL conditions (4.6 vs. 4.Ø, p=.Ø3).

- .The feeling of the group was perceived as more friendly when there was a Human Leader and when there was No Feedback (see Table 5-1).

- The group members seemed more interested when there was no feedback (mean for Feedback= 2.2, vs. 1.7 for No Feedback; p= .006).

Turning to perception of having reached a group consensus, the subjects are correct in reporting relatively high ratings for the HLNF condition. However, they underestimate consensus, relatively speaking, for the NLF condition. (see Table 5-2). When something is as strange and different as a computer-based decision analysis tool, the impressions of subjects as to its helpfulness are not always accurate.

Table 5-1

Perceived Friendliness of the Group, by Condition

Analysis of Variance

	Feedback	No Feedback	Both
Leader	1.7	1.4	1.6
NL	2.0	1.8	1.9
All	1.9	1.6	1.7

Leadership, F=4.96, p=.03
Feedback, F=4.06, p= .05
Leadership x Feedback, F=.10, NS
Group, F=1.62, p=.06

Question:
The feeling of our group was

: 1 : 2 : 3 : 4 : 5 : 6 : 7 :
Friendly Unfriendly

Table 5-2
Perception of Having Reached Consensus, by Condition
Analysis of Variance

	Feedback	No Feedback	Both
Leader	2.7	1.6	2.2
No Leader	3.2	3.0	3.1
Both	3.0	2.3	2.6

ANOVA, Individual Level
Leadership, F=18.97, p=.0001
Feedback, F=10.5, p=.002
Leadership*Feedback, F= 3.92, p=.05
Group, F=5.02, p= .001

Question: Did your group reach a consensus?
: 1 : 2 : 3 : 4 : 5 : 6 : 7 :
Definitely Not at all
Yes

VARIATIONS BY SUBJECT CHARACTERISTICS

The Effect of Age

The older a subject was, the more likely he or she was to have less positive subjective reactions to a computer conference. Most of these relationships are statistically significant; these are shown in Table 5-3.

In the previous chapter, we saw that older subjects objectively perform more poorly. They have fewer typing skills, enter fewer lines, and show less improvement in the quality of their decisions as a result of the group discussion. It is not surprising that the poorer performance is associated with poorer attitudes.

One example of the data underlying the correlations between age and satisfaction is shown in Table 5-4, cross tabulating age by how satisfied the subjects are with their own performance in the group discussion. This is the item that is most highly correlated with age. Note that we unfortunately have very few persons 55 or older; but none of them are highly satisfied with their own performance in this medium.

Table 5-3

Correlations Between Age and Subjective Satisfaction with CC

Item	Pearson's R	p
Problem is interesting- boring	.23	.01
How satisfactory is CC for:		
.Problem Solving	.23	.01
.Persuasion	.18	.05
.Resolving Disagreements	.24	.01
.Getting to know someone	.24	.01
.Exchanging Opinions	.19	.04
Satisfaction with own performance	.28	.01
Agree with Decision	.21	.02
How productive was the group?	.22	.01

Table 5-4

Satisfaction with One's Performance, by Age

Age	1	2	3	4	5-6	N
Under 35	20%	43	21	5	1	56
35-44	12%	31	21	31	5	42
45-54	6%	35	35	12	12	17
55-64	0	0	40%	20	40	5

Chi Square= 46.4, p=.001

gamma= .29

Sex and Subjective Satisfaction

Women tended to rate computerized conferencing higher than men in this experiment, though most of the differences in ratings are not statistically significant. One exception is that the perceived degree to which computerized conferencing is satisfactory for getting to know someone is significantly greater for females than for males (Table 5-5). There is also a statistically significant relation between sex and agreement with the group; the females are more likely to agree with the group (r= -.20, p= .03).

Sex is confounded by typing ability, which is itself related to measures of subjective satisfaction. Women are less likely to be hunt and peck typists (13% vs. 24%) and more likely to consider themselves to be excellent typists (29% of the female subjects vs 7% of the males; p=.01)

Table 5-5
Sex by Satisfaction with Computerized Conferencing for Getting to Know Someone
(1= completely satisfactory, 7= completely unsatisfactory)

Rating	Male	Female
1 or 2	15%	16%
3	11	29
4	17	32
5	27	13
6-7	31	11
Total	100%	100%
N	82	38

Chi square=16.7, p= .01
Point Biserial Correlation= .23

Typing and Subjective Satisfaction

Generally, typing ability is positively related to various measures of subjective satisfaction with computerized conferencing, though most of the relationships are weak and/or insignificant. Exceptions are ratings of the extent to which computerized conferencing is satisfactory for bargaining (gamma=.33, p=.03); for persuasion (gamma= .30, p=.10); for giving and receiving opinions (gamma= .20, p=.03); and the extent to which the group's online conference was perceived as productive (gamma=.15, p=.06).

Effect of Previous Computer Terminal Experience

We have seen that previous experience with computer terminals is related to measures of individual performance, improvement in quality of decision, and the ability of a group to reach consensus. It is also related to some measures of subjective satisfaction, particularly satisfaction with one's own performance in the discussion (see Table 5-6). There are similar, but weaker and not statistically significant relationships with reported perceptions of how pleasant it was to take part in the experiment (p=.15), and the reported friendliness of the group (p=.12).

Table 5-6
Previous Computer Terminal Experience by Satisfaction with One's Performance
(1=completely satisfied, 7= completely unsatisfied)

Experience	1	2	3	4	5-6	N
Never	7%	20	13	27	33	15
Once or twice	7%	33	53	7	0	15
3-10 times	11%	33	22	11	22	18
Frequently	18%	40	21	17	4	72
All	14%	36	24	16	10	120

gamma= -.32
Chi square= 30.8, p=.01

GROUP DIFFERENCES

We have seen in previous chapters that there are pervasive differences associated with group membership, among our "naturally constituted" rather than randomly assigned experimental groups. These differences also occur for subjective satisfaction. Analysis of variance shows that group differences are significant for the following variables, at least at the .05 level:

1. How interesting the problem is perceived to be.
2) How satisfactory the medium is for:
   - Problem solving
   - Bargaining
   - Generating ideas
   - Getting to know someone
   - Exchanging opinions
3) How "friendly" and "productive" the group felt.

SUMMARY

There were some differences among conditions in subjective satisfaction, but they are not very consistent. The Human Leadership condition is associated with improving the process of giving and receiving information, on the one hand, but with making the problem itself seem less clear on the other. Though the medium was rated as more "friendly" with a leader, it was also rated as poorer for "getting to know someone." Feedback was associated with better "giving and receiving information," but also with making it more boring. Thus, none of the structural variations is clearly superior in terms of subjective satisfaction.

There are strong relationships with characteristics of the individual subjects. In particular, older subjects are less satisfied with the medium. There are weak but consistent variations by sex: Women are more satisfied than men. However, this sex difference is confounded by typing ability. The better typists are somewhat more satisfied, and women tend to have better typing skills. Finally, those with previous experience using computer terminals tend to be more satisfied.

There are also significant differences in subjective satisfaction associated with the differences among groups.

If this had been a controlled laboratory experiment with random assignment to groups, we might have seen more correlations between condition and subjective satisfaction variables. However, any such differences are evidently small compared with the overwhelming impact of differences in the characteristics of individuals and in group composition.

CHAPTER SIX

METHODOLOGY: THE AUTOMATED EXPERIMENT

COMPUTER AND HUMAN ROLES IN CONDUCTING THE EXPERIMENT

In the first experiment, we used what might be termed "computer assisted" experimentation for the computerized conferencing condition. All instructions were stored on line, and the computer prompted the experimenter with the instructions to deliver at different points. For the ranking problem, it also checked the ranks entered by each subject to ensure that all items had been ranked once and only once, and prompted for a reranking if an item was missing or used twice. We were quite pleased with the advantages of using the computer as a laboratory tool for group problem solving experiments in this manner, and decided to construct this second experiment as a completely automated one. The computer completely "ran" the experiment, continuously delivering status reports to the experimenter or "monitor," with the exception of allowing the monitor to decide when to actually end the three main phases of the experiment.

Two persons conducted each run. One sat at the monitor terminal and observed the experiment's progress. The monitor had the power to override the automatic progress of the experiment at any point if something went wrong, such as a subject becoming disconnected. The second person circulated from room to room during the training

period, offering assistance. After the training, the doors to the subjects' offices were closed, and the circulating member of the team entered only if the terminal became disconnected, the subject asked for help, or the monitor noticed that something might be wrong.

Initially, the monitor entered the names of the subjects and set the experimental condition. From this point, the experiment proceeded in fifty-seven steps. For instance, step one was the delivery of the initial instructions about how to use a computer terminal to send a comment to the other group members. Progress from one step to the next was programmed on the basis of any of three conditions: completion of a step by a subject, the passage of a certain number of minutes, or completion of a step by the entire group. For instance, step two was the entry of three practice comments by each subject. As they finished the third comment and received any waiting items, they were then automatically given the second set of instructions, consisting of a rank ordering instruction and the first practice problem. (See the Appendix for the text of this instruction, which was "step three," and for the full text of all other instructions). Thus, the subjects were able to proceed through the training at their own pace.

An example of a step that was executed as a function of the completion of an operation by all five subjects was the delivery of the first table showing the rankings on the practice problem. An example of a time-determined step, with an override possible by the experimenter, was the delivery of the sixty-minute warning half-way

through the arctic problem. When all five subjects had completed their initial rankings, a timer was set and the discussion guidelines delivered to them simultaneously, so that they all began the problem discussion at the same time. The sixty-minute warning could have been sent automatically. However, there were circumstances in which "clock time" on the computer in Newark, New Jersey was not identical with the effective time on line for the subjects. For instance, the local Telenet node could have gone down, keeping the subjects incommunicado for some time, or an individual could become disconnected or have a paper jam and lose time until the problem was corrected. When receiving the warning notice, the monitor decided, based on whether there had been local problems, to deliver the warning to the subjects then or wait so as to permit sixty minutes of real discussion time, rather than purely clock time.

Some progressions to a "step" could be determined on the basis of a combination of criteria. For example, the algorithm for the delivery of a new table (or two tables, for the feedback condition) showing the groups ranking was the following:

1. When a table was printed, a timer was set. Even if there were no subsequent rerankings, a new table was printed after twenty minutes to make sure that the group was aware of its status.

2. When an individual reranked, a timer was set for ten minutes, during which time any additional rerankings were collected. Then a new table was printed, incorporating all the changes. The timer set

by a reranking operation overrode the elapsed time criterion.

The computer was also used to completely "block out" the remainder of the EIES system. Four simple commands were provided, in place of the usual myriad of possible choices available. For example, when entering a comment, one is usually asked to make several choices: whether to give the comment a "key" or title, whether it is "associated" with any previous comment, and whether the author wishes to sign it or use a pen name or anonymity. The "+ENTER" command given the subjects skipped these choices and entered the comments automatically, without keys or associations, and with a regular signature.

Only these commands operated during the experiment. If another command was given that would be normally valid and that could, for instance, take them to the message system or to another conference, they were told that this was an invalid command and asked to try again.

One of the experimental features was three "gates," where those who had completed a step were held and blocked from further communication with the group until all had reached the same "gate" and were simultaneously "let out." One of these was at the completion of the initial individual ranking for the arctic problem. The terminal simply would not accept any text entry until all five subjects had completed their rankings and received their discussion guideline instructions and the table of rankings for the whole group. As with

other instructions, this was programmed to be as polite as possible. Each subject was asked to please wait for the others to finish their ranking and be ready for discussion before entering anything further. As each individual completed the ranking, the others were kept informed of this progress. These one-line status reports looked like:

JANE DOE (JANE,9Ø1) is ready to begin discussion.

We found that without these "status reports," the subjects felt frustrated and wondered if "the machine was broken." With them, they felt informed about what was happening.

The other two "gates" were after the second practice problem and before the the final group ranking for the arctic problem. At the "lunch break," progress to the next step (printing the arctic problem on each terminal) occurred only when triggered by the monitor, who first checked to make sure that there was sufficient paper on each terminal to last the afternoon.

Taking the Experiment Into the Field

Since the experimental procedures could be accessed by anyone in any location with a telephone line and a computer terminal, we had constructed what might be termed a "laboratory without walls." We were able, by transporting portable terminals, to bring the experiment to staff groups in their offices around the country. These managers and professionals would not have been willing to

travel to a laboratory, but they were happy to have the experiment, termed a "participatory seminar," brought to them. As a quid-pro-quo to the sponsoring organizations, a free seminar, open to anyone invited by the sponsor, was presented at the end of the experiments at each location.

There were some inevitable technical problems. Though we brought seven terminals (one extra), sometimes more than one terminal burned out before the end of the day, in which case we gave up the monitor terminal and lost the data for that group. Sometimes telephones were located nowhere near electric power outlets in offices and we had to string long extension cords. Sometimes the office phones had "noise" on the line, and we had to move participants to a better line. Generally, with seven terminals plus a large case of paper and forms being carried by two persons, we felt a bit like pack mules or a travelling circus. However, with at least an hour's set-up time, the travelling road show was able to successfully "go on" in most locations.

Training and Monitoring Aids

Without the use of the computer, we would have needed an assistant with each subject during the training and at other points, to offer help when needed. In addition to intruding on their privacy and possibly adversely affecting the "natural" progress of the discussion, this would have been expensive, since a large number of people would have had to be transported to each site.

During the training, the computer checked that each subject had correctly mastered each of the commands, using a form of computer-assisted instruction. For example, to test understanding of the reranking instruction, "+ORDER," the subjects were asked to move the items which were third on their lists to the first position, and leave everything else in the same order. This request was individually tailored for each subject, based on the initial order. For instance, if the subject had entered "B Mousse" as the third ranked item, the instruction was to "Move B Mousse to become the FIRST item." If this was performed correctly, the computer confirmed it with "That was correct, very good." However, if it was reordered incorrrectly, the computer responded, "Sorry, that is not correct. Please try again." The monitor was also informed that there had been an error. If the subject incorrectly entered the new order a second time, the computer showed the subject what the correct entry would be. Meanwhile, the roving assistant, alerted by the message on the monitor terminal, would offer further explanation if necessary.

The monitor frequently used the "+STATUS" command, also available to the subjects, to receive a report on whether each person was on or off line, and the last comment read. If a subject was off line, the assistant (literally) ran to reconnect the terminal. If a subject lagged far behind the rest of the group in the discussion, the assistant checked to see if there was a problem.

The monitor had a number of special commands to keep track of the

proceedings. For example, "+STATES" showed the location of each subject in the experiment at any point in time. Another command allowed the monitor to reset the subject to another step if there was a problem. For example, after entering an initial ranking for the arctic, some subjects wished to change a a mistaken entry before the ranking was shown to the other group members. The monitor could then set the subject back to the initial ranking step.

Problems: Automated Errors

The problem with a programmed process is that one must specify in advance all of the contingencies and "go to" operations. Of course, it is not possible to anticipate all of them in advance of running an experiment, or even with a limited number of trial runs and subsequent adjustments to the software, such as we used. One example is that we had decided to make reranking easy for subjects by enabling them to type in a partial reordering and then doing a carriage return, which meant "leave everything else the same." This worked well in the pretests, which were conducted on local lines from Upsala or on government tie lines. It produced some errors when using TELENET, which sometimes generated spurious carriage return signals as a form of "noise" on line, entering an order not intended by the subject. If this Telenet-generated carriage return occurred before the subject typed in the initial ranking, the original alphabetical listing appeared as the rank order. It was not actually entered unless the subject confirmed it as correct, but confused subjects sometimes confirmed the accidental alphabetical listing.

The mistake then became clear, and the subjects always informed us when this happened, and were told to reenter the order correctly. We thought all was well until a final check on the results of our analyses, including a detailed check of every item of data that had been used. We discovered that our automated analysis program (see below) had picked up the first "initial ranking" and used it in computing the Kendall's coefficient for initial pre-discussion agreement, rather than the corrected pre-discussion ranking, when mistakes had occurred. When we specified the program, we had not anticipated this contingency. Therefore, our initial sets of analyses, including some that had been published, were slightly wrong (eight cases of 120 had some incorrect data; not enough to change the general nature of the findings, but enough to change the specific numerical results of the analyses). Thus the end result of our automated analysis and an unanticipated technical flaw was the temporary creation of some incorrect results.

Automated Analysis

A complete log and transcript were kept for each experiment, showing the time and results of any reranking operation by any subject, and the time and length of all comments entered. Computer programs were used to automatically analyze much of this information, including:

1. The number and percentage of lines and comments entered by each subject;

2. The Kendall's coefficient at any point in time;

3. The deviation scores (from criterion) for the initial, last subject reranking, group ranking, and final individual (post-discussion) rankings.

This saved some labor and should have reduced errors by obviating the necessity to re-key data in order to analyze it. Unfortunately, there was a mistake in the routine which switched labels among the various Kendall's coefficients. This was not discovered until after an analysis had been completed and some initial results had been released, with an incorrect label on the tables.

A more valuable and trouble-free procedure was using the computational power of the computer in "real time" to provide "decision support" and "experiment support" calculations and displays that would not be simple to do manually in real time, without slowing the progress of the experiment or decision making process. For instance, it would be conceivable for a human with a calculator or a separate computer to enter ranking data and compute the average ranks and coefficients of agreement that were provided in the "feedback tables"; however, this would noticeably slow down the flow of the group process.

A SIMILAR AUTOMATED EXPERIMENT

In addition to our own work, EIES has been successfully used by other investigators for a completely automated experiment comparing recall

of communications with actual communications on line (see Bernard, Killworth, and Sailor, 1979; the software was developed by Peter and Trudy Johnson-Lenz). The particular use made of the computer was quite different than that for our experiment, and can help to illustrate the possibilities made available by the technology.

In our experiment, it was the "treatment" itself that was complex and which relied upon the computer to take the subjects through the many steps of a synchronous experiment in which the specifics were contingent upon the condition. In the experiment on informant accuracy in recalling communications, there was basically only one treatment, an interview administered by computer. However the "communications window" varied; there were 37 windows representing different combinations of "lag" and "width." Width is the amount of time over which informants were requested to report their behavior, and ranged from one to thirty days. "Lag," the amount of time elapsed since the end of the window, varied from one day or even less to sixty days.

The computer was used to schedule interviews and to administer them at a time convenient to the volunteer subjects. When a subject signed on, the computer determined if it was "time" for another interview, based on calculations related to the number of interviews completed by the respondents (each took up to 37, one for each window), relative to the progress by other subjects and the total time available for the completion of the study. If it was "time," the computer randomly selected a "window"; these random selections

were based on windows completed not only by the respondent by also by the totality of subjects, so as to keep even coverage of all the windows in the experiment.

Over the period of the four months that the experiment was conducted, the computer also kept track of the actual communications of each subject for each window for which an interview was collected. In addition, features designed to meet the needs of subjects kept the experimental procedure sufficiently flexible so that the subjects could tolerate such a long-term study. A subject, when informed that it was time for another interview, could take a "rain check" on the interview, postponing it until the next time he or she signed on line. Only one rain check was allowed, however; the subject could not use the system for communication on the subsequent sign-in until the interview was completed. A second programmed condition providing some flexibility was a "harrassment limit"; each individual set a time for interview length beyond which he or she was unwilling to go. If the subject was not near his or her own "harrassment limit" after completing an initial set of questions, a second set was administered; if the harrassment limit was near, the computer did not begin administering the second set of questions. Most subjects picked a harassment limit near twenty minutes.

A third feature of the experiment helped to make it more interesting for the subjects and to provide motivation beyond the modest sum they were paid for participation. The subjects could check on their own accuracy of recall by using a routine called "feedback." This showed

the subject the actual communications data matching the subjective reports supplied for a completed interview.

CONCLUSION

Despite some problems with automated errors which it took us over a year to completely identify and correct, we continue to be favorably impressed with the use of a computerized conferencing system as a tool for the experimental study of human group communication. For both our own study and that by Bernard et al., the use of the computer made possible more complete data collection on subject behavior than would otherwise have been possible. Furthermore, using computer assistance or automation, it is possible to much more closely replicate most manipulations and variables used in a previous experiment to introduce variations designed to extend the findings, as we did in repeating the use of the arctic problem and instructions in the second experiment. This would be termed a variety of "constructive" replication according to the taxonomy developed by Kelly, Chase and Tucker (1979), who point out contributions which replications can make to the generalizability of previously reported results.

In sum, we believe that systems such as EIES offer opportunities for future investigators to use automated experiments to study larger groups, over longer periods of time, with more complex experimental designs and treatments, and more complete data collection than has previously been possible.

CHAPTER SEVEN

SUMMARY AND CONCLUSIONS

SUMMARY OF FINDINGS

The purpose of this experiment was to assess the effectiveness of alternative communication structures within a computerized conference to support group decision making among managers and professionals. Listed below are our initial hypotheses, and the corresponding findings. Groups composed of managers and professionals within a variety of organizations were given a 15-item ranking task with a "correct" or criterion solution. Their task was to reach agreement on the "best" rank order within two hours, using a specially constructed version of EIES (the Electronic Information Exchange System). Two alternative means of structuring the conferences were employed, in a two-by-two factorial design. Groups with "Human Leadership" elected one of their members to lead the group in its decision making discussion. Groups with "Computer Feedback" were given periodic tables which displayed the current "group decision" in terms of the mean rankings of items, and the degree of consensus about each of these items.

Decision Quality and Degree of Consensus

Initial hypotheses and summary of findings:

1) Human leadership will improve amount of consensus (some support)

2) Human leadership will improve quality of decision (not supported)

3) Computer feedback will improve amount of consensus (some support)

4) Computer feedback will improve quality of decision (On the contrary, some indication of negative impact)

5) There will be interaction between human leadership and computer feedback (some support, for consensus)

The word "some" is used in summarizing the findings, because it depended upon how the dependent variables of quality of decision and consensus. We had three different measures of each of these variables. In each case, the findings were statistically significant only for one of the three measures.

We found that when differences in group composition were taken into account, there were no significant differences either in the absolute quality of the group decision or in "percentage improvement". Groups in all conditions made substantial improvement over average individual decisions, following discussion.

"Collective intelligence" was defined as the ability of the group to make a better decision than could have been made by its "best" member without discussion. This occurred for half of all the groups. Those groups with Computer Feedback were significantly less likely to achieve collective intelligence.

For those groups with a Human Leader, the knowledgeability of that leader greatly affected the quality of the group decision.

Turning to ability of the group to reach consensus, we found high

group consensus in all conditions for the final, post-discussion reporting of a group decision (mean Kendall's coefficient of concordance of .972). There were no significant differences. However, there were significant differences in the amount of agreement among the final individual rankings which occurred during the discussion itself. The HLNF and NLF conditions were clearly superior. In other words, there was a significant interaction; either aid helped, but in combination they conflicted and were not helpful for reaching consensus.

Group Process

Hypotheses:

6) Human leadership and computer feedback will affect the process of communication as follows:

a) There will be more re-ranking with computer feedback.

b) There will be more discussion with human leadership.

There is a small but significant tendency for less discussion with feedback tables. With a human leader, there is less reranking, but the presence of the feedback tables has no effect on amount of reranking. Thus, though our initial hypotheses were along the right lines, we stated the cause and effect incorrectly. It is not for instance, that there is "more reranking with computer feedback," but

rather, comparatively speaking, there is "less" with human leadership.

Though the experimental variations in structure produced these observable differences in group process, this had no significance for group performance. Neither amount of discussion nor frequency of reranking were related, on the average, to group consensus or quality of decision.

7) There will be more inequality of participation with human leadership, with the leader more likely to dominate the discussion.

There was no association between condition and the likelihood of dominance. Only one out of the six groups in each condition had a dominant individual. The Human Leaders, when present, did tend to contribute slightly more to the discussion, but not enough to come anywhere near "dominating" the discussion in terms of volume of communication.

Subjective Satisfaction

Hypothesis 8: Human leadership will be associated with greater subjective satisfaction than computer feedback.

Findings: There is no consistent difference among conditions in subjective satisfaction. However, there are significant variations associated with differences among individuals and groups (see below).

Variations Associated with Subject Characteristics

Hypothesis 9) Typing speed will be positively associated with quality of decision and ability to reach consensus. (supported)

10) Previous computer experience will be positively associated with quality of decision and ability to reach consensus (supported).

We found that both typing skills and previous experience with computers are positively related to improvement in quality of decision, the ability of a group to reach consensus, the amount of participation in the discussion, and subjective satisfaction.

11) Age will be negatively related to quality of decision, ability to reach consensus, lines entered, and subjective satisfaction (supported).

Older subjects performed more poorly and had more negative attitudes. They contributed fewer lines and improved their rankings less as a result of discussion. Groups with older members were less likely to reach consensus. Older participants had consistently more negative attitudes, including feeling much less satisfied with their own performance. This set of findings has serious consequences for penetration of the medium into managerial decision making processes, since most senior executives are older.

12) Females will be more satisfied with the medium than males (supported).

Sex composition of the group had much more pervasive influence that we had hypothesized. There was a tendency for groups with more females to improve their decisions more. Females contributed more to the discussion than males, on the average. They also tended to be more satisfied with the medium than males, though most of the differences are not statistically significant. Of course, sex differences are confounded by differences in typing ability. The females had better typing skills, and we do not have enough female subjects at all levels of typing skill to separate the effects of sex and typing.

The Pervasive Influence of Group Differences

A field experiment employing actual groups in their usual setting has the advantage of being more realistic and more generalizable to "real life" use of the medium than a controlled laboratory experiment with randomly (artificially) constituted groups of subjects. However, the field experiment design suffers from the analytical difficulty that differences among subjects and among groups may be confounded by differences in the experimental treatment (as they were for this study), and "drown out" the effects of the "treatment." If we had used the laboratory experiment model, we may have found more statistically significant differences related to the use of computer

feedback and/or human leadership in a computer conference. However, if in "real life" such differences due to structure are small compared to the overwhelmingly powerful effect of differences among individuals participants and groups, perhaps it is best to have discovered the relative explanatory power as part of the experimental design.

We do not wish to return to the artificiality of using college students or other subjects who are compliant but not representative of the managers and professionals for whom we are attempting to build group decision-support tools. Thus our decision for the design of the third and final experiment in this series was to find our subjects among the employees of a single organization, so that even though the experiments were run "on-site," we could control assignment to group and have a more homogeneous set of subjects. At the time of the writing of this report, we have conducted the final experiment. It uses middle-level managerial and staff employees of one of the hundred largest corporations, and examines the effect of "pen names" on the process and outcome of risk-taking group decisions (See Hiltz, Turoff, and Johnson, forthcoming).

NOTES ON STRUCTURE

The "structure" of a computer-mediated communication system refers to the many design choices that have been made which will affect the nature and flow of communications within a group.

For example, one can think about an ideal structure for synchronous (real time) group communication. This is different than the ideal structure for an asynchronous conference. It is very important for subjects to keep "current" in such circumstances, even though they may spend several minutes composing an entry. There must be a way to "interrupt" them with priority information, even though they are in composition mode. For this experiment, we made an arbitrary decision, based on observations during our previous experiments and during pretests for this one, that it was crucial that a one-line "interrupt" be broadcast to all members whenever a group member changed rank orders. In the normal EIES mode of operation, it would be up to a user to decide when and if such an interrupt should be sent. Since we provided this immediate and automatic notification to all groups, we cannot measure the extent to which it was indeed helpful. However, we do feel that it is one of the factors which enabled the groups in this experiment to reach such high levels of agreement within the time limit.

Whereas our subjects had a single screen and could EITHER send or receive at any time, our observations indicate that it would probably be better to structure the flow so that sending and receiving are two separate streams, and may occur simultaneously. By having a large display terminal and another printer working independently, communications being composed could appear on the screen, and simultaneously, communications being sent by other group members would be printed. Participants could thereby pause to read incoming communications without have to complete or abort the sending of their

own communication.

We structured human leadership in the simplest manner possible. Group members elected a leader, and only normative pressure (no software features) supported this leader. One can imagine many other ways of structuring leadership. For instance, a computer analysis could be used to identify the group leader during the training and practice session according to which person had a communication profile which best matched that of successful leaders in this medium in the past. In this experiment, those groups which had a leader within their organizational context tended to elect that person, even though the person with highest rank might not have had the skills to be effective in this medium. Only those composed of peers seemed to feel free to select on the basis of performance during the practice. So perhaps "computer appointed" leaders would have been more effective.

"Leadership" might also be supported by software by permitting only the group leader to have certain powers, such as calling for a vote or viewing the results of an analysis of the group choices. In this experiment, anybody could "vote" or rerank at any time, and all participants received the same decision aid display.

Pen names or anonymity offer interesting variations. In pure anonymity, there is no individualizing information on any communications whatsoever. With a pen name, each participant's contributions are uniquely identified and can be responded to,

without revealing actual identities. For example, participants might be identified by numbers ("one," "two," etc.), colors ("red," "blue," etc.) or by purely hypothetical names they choose, such as "The Monster" or "Julius Caesar." Anonymity or pen names might be prohibited entirely (not allowed as an option); permitted as an option in addition to "real" signatures on entries, or required by having items entered this way automatically. The pen names or anonymity might relate to text communications, votes, or both.

The point is that there are many variations in structure that can be created. We found some variation due to the structures we provided for this experiment. Perhaps stronger variations would have occurred if we had implemented the structures differently.

Independently of variations in the structure of a computer-mediated communication system, one can vary the implementation. This includes training procedures, interface, response time, etc. As compared to the first experiment, we gave subjects in this study a longer training time, plus actual practice with the type of problem they would be asked to solve. We believe that this is one of the reasons why, as compared to computerized conferencing groups dealing with the arctic problem in the first experiment, these groups reached higher levels of consensus, improved their decisions more, and had higher levels of subjective satisfaction.

IMPLICATIONS FOR DECISION SUPPORT SYSTEMS

The generally accepted objective of a decision support system (DSS) is to interface a manager's judgement and a set of appropriate models and data bases which directly relate to a problem and which provide aid in reaching decisions. Keen and Morton (1978) express this in terms of problems that can be organized so as to be "semi-structured":

> The second level, of semi-structured tasks, is where DSS can be the most effective. These are decisions where managerial judgement alone will not be adequate, perhaps because of the size of the problem or the computational complexity and precision needed to solve it. On the other hand, the model or data alone are also inadequate because the solution involves some judgement and subjective analysis. Under these conditions the manager plus the system can provide a more effective solution than either alone (p. 86).

Although this is a rational view of DSS in current practice, it is unnecessarily confining. Our concern here is not with what DSS have been, but with what they could be. Until now DSS have involved a single person interacting with data bases, models, and analytic routines. We believe that if it were embedded within a computerized conferencing system (CCS), DSS could be a general tool for the support of GROUP communication and decision making. Our colleague Julian Scher (1981) refers to this concept as "DDSS": Distributed Decision Support Systems.

DDSS and the Structure of Organizations: Some Assertions

The current trend in DSS is to move problems from ill-structured to semi-structured and, ultimately, to well-structured situations. As Simon observed, computers facilitate centralized control. The more structure, the more centralized control is possible. What computers achieve in organizations was suggested by von Bertalanffy (1968): the computer, by imposing a structure on information flow between segments of an organization, causes progressive "mechanization and specialization" of the work of the segments. This reduces interaction and increases inequality between segments, which in turn leads to centralized decision making.

Traditional computer systems (Information Systems and Decision Support Systems) also promote formalized interactions between segments and usually require those interactions to be concise, quantitative forms of information transfer. Very specific inputs constrained to the formats of the system are required and very specific outputs are generated. Although this leads to efficient operation of the organization under regular or stable conditions, it does have negative consequences. As Mowshowitz (1976) stated,

> The efficacy of hierarchical organization is intimately linked to goal structure. If the sole purpose of an organization is productive efficiency, then hierarchical structure may be warranted. But the subordination of individual aims required to secure this objective cannot be achieved without cost. Is there any reason to believe, for example, that reduced

> information transmission between individuals in different units of an organization is inherently desirable? In the short run, one might anticipate certain savings in time and effort. However, the long-term consequences of diminished interaction are likely to show up as a kind of "genetic impoverishment" similar to that observed in populations with excessive inbreeding (p. 79).

As organizations become more specialized and centralized, they cannot easily adapt to a changing environment; thus they suffer from a lack of "resiliency" in the ecological sense. To date, the impact of the computer on organizations has been largely to establish models and data bases which describe the organization at a particular point in time. With the passage of time these models become templates which prescribe the organization or constrain it to behaving like the abstraction contained in the computer system. The only way to counter this trend over the long term is to ensure that these structures are changed as fast as the environment changes. One solution is to provide communication processes that will allow for change. Computerized conferencing technology to do this exists. These systems are also likely to increase information transmission and decentralization. The problem in adopting them lies not so much in the computer and information systems currently in place. Rather, it lies in our lack of faith in these systems, and/or an inability to act because of the segmentation that has already taken place, even at higher levels in many organizations.

CONCLUSIONS

Most Decision Support Systems use computers to support interaction between individuals and a structured model, analytic routine or a data base. However, many problems are unstructured or at best semi-structured, and are dealt with by groups of managers within organizations. When dealing with nonroutine problems, the decision-making groups are often geographically and organizationally dispersed. Thus a decision support system for these groups must include communications, structured to support the decision-making process, among members of the group.

Our experiments indicate that computerized conferences can effectively support group communication and decision making. This is particularly true when they are structured to provide aids suitable to the problem at hand, such as explicit leadership roles or data display and analysis of options being considered by the group.

For this study, with group size of only five, human leadership was more effective than computer feedback. For very large groups (20 or more), we suspect that computer feedback (analysis and display of data related to the group decision) would prove more valuable than it did in this experiment.

Though there were some effects of experimental variations in structure or "groupware," social context variables relating to individuals and group attributes were more powerful determinants of performance. Previous experience with computers, typing ability, age, and sex all affected individual performance. On the group level, the knowledgeability of the participants and particularly of the leader, if one was elected, were crucial. There were also noticeable differences in how well the groups were able to work together on line, probably as a result of previously formed social relationships.

Thus, we must conclude that some groups are simply much better candidates than others for using computerized conferences for discussion and decision-making. Groups composed of participants with some previous experience using computer terminals and groups with cooperative rather than competitive social histories are recommended. On the basis of the clearly superior performance of the subjects in this experiment as compared to those in the first experiment, we would also stress the apparent importance of adequate training and practice with this medium before being asked to use it to solve a difficult problem, and of adequate time to complete the task, which is likely to be a longer elapsed "clock time" than would be necessary for a face-to-face meeting.

APPENDIX I: TRAINING AND PRACTICE INSTRUCTIONS

A. Initial Instructions

Hi! Today you are going to learn to use a computer mediated system for human communication. We are going to teach you how to "talk" with the other members of this conference, by typing what you want to say on this terminal and having it sent to the other conference members. Then we are going to teach you a special set of commands to enable you to rank order lists of items, since that is the type of problem your group will have to solve after you have practiced using the system.

First, we want to show you how easy it is to type on this terminal.

HOW TO TYPE ON THIS COMPUTER TERMINAL

There is room for a certain number of spaces on a line. The spaces are marked on a strip just in front of the print mechanism. You can always look and see how far you have typed on a line. When you press the RETURN key, the carriage will return and give you a new line.

PLEASE DO NOT TYPE PAST THE ARROW ON YOUR TERMINAL BEFORE PRESSING THE RETURN KEY

To make a blank space, you press the large space bar on the bottom.

The letters on this terminal are just like a typewriter. To type a capital letter or a character in the upper case range, hold down the SHIFT key -- you will find one of these on the left and one on the right. The numbers are all on the top row, which is also like a typewriter. However, there are some ways in which typing on this terminal differs from a typewriter.

1. Typing in a "SCRATCHPAD"

When you want to say something to the other conference members, you will be typing what you want to say into what is called a "SCRATCHPAD". These are numbered lines into which you type the text of what you want to say. The terminal will tell you when it is ready for you to start typing by printing

ENTERING SCRATCHPAD:
1?

You can now type the first line of what you want to say on this line that begins with a 1? When you are finished typing a line, press the RETURN key. This will give you a new numbered line which looks like

2?

When you have typed what you wish on line 2, and need more lines, pressing the RETURN key at the end of every line will give you a new numbered line on which to type. ALWAYS WAIT FOR A QUESTION MARK TO APPEAR BEFORE YOU RESUME TYPING. Even if what you have to say

takes only one line or letter, always press the RETURN key after you have typed a line. Pressing the RETURN key enters what you have typed into the computer. Until you press the RETURN key, nothing can be done with the line you have typed.

Sometimes, the computer will stop in the middle of printing things, and will not give you a question mark (the signal that you may type something in). Just be patient. It is finding something else to deliver to you. When it has delivered everything that is supposed to come to you, it will give you a line number or a question with a question mark, and then you can type in again.

2. Canceling a line

Since what you type does not go to the computer until you press the RETURN key, you can change your mind or correct a mistake before sending it. Most people do not bother to correct minor typing errors, as long as the meaning is clear. However, if you want to cancel a line and retype it, hold down SIMULTANEOUSLY the CONTROL (CTRL) key and the X key (think of it as drawing a big X through the line you have started to type, and starting over again. This is the one time when you do no need to wait for a question mark).

HOW TO SEND WHAT YOU HAVE TYPED TO THE OTHER CONFERENCE MEMBERS

Once you have typed into your scratchpad what you want to say, you can send it to the other members of the conference by typing

+enter

as the first and only thing in a new line of your scratchpad, and then pressing the RETURN key.

What you have typed will now be sent by the computer to ALL of the members as a conference COMMENT.

The +enter is a command which must be entered precisely. The + must be the first character on a new line. There can be no space between the + and the enter. It must be followed by a carriage return.

Whenever you ENTER a comment, you will automatically receive waiting comments that have been entered. YOU MUST KEEP TYPING THINGS IN AND ENTERING THEM, IN ORDER TO KEEP RECEIVING COMMENTS FROM THE OTHERS.

You will also receive a copy of your entered comment, so you can see what it looked like. A conference builds up a common transcript of all of the comments entered by the members, and each of the comments entered by you and the other members is given a number.

3. THE +STATUS COMMAND

If you want to see which other members of the conference have read a specific comment at a specific time, type

+status

as the FIRST AND ONLY ENTRY ON A NEW LINE IN YOUR SCRATCHPAD, and press the RETURN key.

This will give a list of the last comment number received by each member of your group.

SOME IMPORTANT THINGS YOU MUST KNOW

1. The system may ask you some questions.

Type y and press the RETURN key for YES.

Type n and press the RETURN key for NO

2. If you want to look at what you have typed, you may roll the paper up. However,

PLEASE DO NOT TRY TO ROLL THE PAPER BACK DOWN

or it may jam. The computer automatically continues on the same line, even though you have moved the paper. You may roll up the paper at anytime you wish, as long as the terminal is not printing. This will not effect what you type.

3. In addition to the other members of this conference, there is a Monitor whose number is 912. The Monitor will occasionally send you instructions asking you to do certain things.

4. If by any chance you get an unexpected question, and think you may be out of this conference by mistake, type

+xpt

and press the RETURN key as the answer to that question. That will get you back into this conference.

YOUR FIRST PRACTICE

PLEASE DO EACH OF THE FOLLOWING WHEN THE TERMINAL PRINTS

ENTERING SCRATCHPAD:

1?

a) Type in a greeting or comment to the other participants, that is one line in length. Then press the RETURN key. The terminal will print

2?

b) In typing the second line of your initial message to the others, type in one or two words, and then try canceling it by holding down the CONTROL (CTRL) key and pressing X at the same time. The terminal will repeat ? and you type in the line again.

c) Add another line or two if you like to complete your first comment to the group. Then type

+enter

as the FIRST AND ONLY THING ON A NEW LINE IN YOUR SCRATCHPAD, and press the RETURN key.

What you have typed has now been sent to all members of the conference as a conference COMMENT. You have now entered your first COMMENT into a computer conference!

d) Continue chatting with other members of the conference until you receive your first practice problem. Use the +status command once or twice in order to see where others are in the discussion.

PLEASE TEAR OFF THESE INSTRUCTIONS AND REREAD THEM BEFORE TRYING YOUR FIRST PRACTICE

B. Second Instruction and Practice Problem
RANK ORDERING

You all seem to be doing very well.

Now, we are going to teach you some more commands to enable you to enter, display, and change rank orders of items.

Here is your first practice problem.

THE DELICIOUS CHOICE

You have arrived at your meeting a bit hungry, and your host has offered to make a dessert for all of you, if you can agree on a single choice.

Please enter your rank order for the following five choices, when the computer asks

Letters in rank ORDER?

The five choices are:

A. CREPES Suzettes
B. Chocolate MOUSSE
C. Apple PIE
D. Black Forest CAKE
E. STRAWBERRY Shortcake

You enter the order by typing in the letters corresponding to the items.

Thus, if you typed cdeba and pressed the RETURN key when asked "Letters in rank ORDER?" as follows,

Letters in rank ORDER?cdeba

you would create a rank order of:

1. C. Apple PIE
2. D. Black Forest CAKE
3. E. STRAWBERRY Shortcake
4. B. Chocolate MOUSSE
5. A. CREPES Suzettes

Here are the items that you are to rank:

C. Table Explanation- All Conditions
TABLE of All the RANK ORDERS

Periodically, the system will compile a table of all the rank orders currently entered by each person in your group, so that you can see how close you are to consensus. The first table will be printed out for you when all members have completed their initial orders.

Additional Table Explanation- Feedback Conditions
THE GROUP CONSENSUS TABLE

The final display you have available is a table that shows what the group decision would be at this point if all the rankings were averaged. It also shows how much agreement there is on each item at the present time.

Agreement reaches 100% if all group members assign the same rank. It would be 0% if half ranked it at the top(#1) and half ranked it at the bottom (#5).

You will also receive an example of the group consensus table that will be compiled and printed for you from time to time, based on your initial orderings in "The Delicious Choice".

D. ORDER Command Instruction for Reranking
THE +ORDER COMMAND

You will need to change your listed order so that the group can reach agreement on a common order. Here is how you do it.

Whenever you have decided, based on the discussion and looking at the TABLE of current orders, that you are ready to change your order, type

+order

as the FIRST AND ONLY THING ON A NEW LINE IN YOUR SCRATCHPAD, and press the RETURN key.

This will list your current order.

Then it will ask,

Letters in rank ORDER?

Type in all the letters in the desired new order, all in a row. If, for example, your NEW order is going to be C D E B A, you would type in cdeba as follows:

Letters in rank ORDER?cdeba

When you use +order, the computer will begin compiling a new table to enter into the conference and show the others the changes you have made. This table will be entered for all to see about ten minutes after any person uses +order. As soon as you complete a +order, a one line statement of your new order will be sent to all others.

Now practice this way of changing your ranking to move the item that you have ranked THIRD on your list to be FIRST.

E. Shortcut Form of Order Command
MORE ABOUT THE +ORDER COMMAND

You have learned how to change your listed order by typing +order and simply typing in the letters of your NEW order when the system asks

Letters in rank ORDER?

However, if you want to change the position of only one or a few items, you need not type in all of the letters again. You can simply type in the letters of the items that you want to change. Let us say that your ranking of the dessert items was

A. CREPES
B. MOUSSE
C. PIE
D. CAKE
E. STRAWBERRY

and you now wanted to place B MOUSSE after D CAKE. You would simply type

+order

as the first and only entry on a new line in your scratch pad, and press the RETURN key. When the system asks

Letters in rank ORDER?

you would type in db, as follows:

Letters in rank ORDER?db

and you would have thus very easily created the NEW order of

A. CREPES
C. PIE
D. CAKE
B. MOUSSE
E. STRAWBERRY

This simple way of using +order goes to the location of the item whose letter you typed in first, and puts the items whose letter or letters you typed in next immediately after this first item. All unlisted letters stay where they are.

Here is another shortcut

If you type +order db

(That is +order followed by a space, followed by letters)

The computer will skip printing out your current order, and just make the change indicated. It will then show you the new order and ask if it is correct (what you intended.)

Please try this simple way of changing your order by putting the item that is now FIRST on your list back to be THIRD on your list.

NOTE: On this and the previous order practice, the computer checks to see if the requested command was entered correctly. If correct, it says "good" and goes on to next instruction. If incorrect, it explains what was wrong and asks the subject to redo it correctly.

F. Instruction to Complete Practice Problem

Now, please use the +enter to discuss your rankings of desserts, and +order to change your rankings, until your group has reached a unanimous decision on your first delicious choice.

Each time somebody changes their order with a +order command, the group will receive an updated table five to ten minutes later.

NOTE: Monitor ended the dessert practice problem when the group reached agreement on the first choice or when lunchtime approached, whichever occured first.

G. Final Practice Problem- No Leader Conditions

Here is a final problem for you to practice on. Please rank order the five potential Presidential candidates listed below in terms of your perceptions at this point of how effective a President they would be.

We want you to practice the initial ordering one more time. No table showing your ranking of candidates will be printed, since your ranking of the candidates is confidential.

Please enter your ordering of the candidates as a series of letters that corresponds to the following names:

H. Final Practice Problem- Leadership Condition

Here is a final problem for you to practice on. Please rank order the five members of this group in terms of your perceptions at this point of how effective they would be in leading a discussion. We want you to practice the initial ordering one more time. No table showing your responses will be printed, since your ranking of the group members is confidential.

Please enter your ordering of the group members' leadership ability as a series of letters that corresponds to group members as follows:

I. Break Instruction

NOTE: This was printed out on each terminal when the final ranking practice was completed. Then the subjects gathered for lunch and review of the "Crib Sheet" and discussion of any questions or problems pertaining to the practice session.

If you have any questions or comments, please ask an assistant.

We will have a break now. Please do not enter anything more on the keyboard.

After lunch, the problem was waiting for each subject, printed out on his or her terminal, with instructions to rank order the importance of the fifteen items by entering the letters corresponding to the items. Each person was informed as each of the others completed the ranking and was ready for discussion. When all five had completed their initial ranking, one (for no feedback) or two (for feedback conditions) was printed showing the rank orders of the five participants.

Then the discussion instruction was received.

J. Begin Disscussion Instruction

All members of the group have now completed their initial ranking.

You may begin your group discussion and attempt to reach consensus on the ranking. Remember that to enter a comment to the group, type +enter as the only entry on a new line of the scratchpad, and press the RETURN key. To change your rank order, use +order.

You will have up to two hours in which to complete discussion and reach consensus. You will receive a warning at the end of 60 minutes and 90 minutes.

At the end of the discussion, you will be asked to report the rankings agreed upon by the group.

DON'TS

1) Do not make early, quick, easy agreements and compromises. They are often based on erroneous assumptions that need to be challenged.

2) Do not compete internally. In this situation either the group wins or no one wins.

DO'S

1) Pay attention to what others have to say. This is the most distinguishing characteristic of successful groups.

2) Try to get underlying assumptions regarding the situation out into the open where they can be discussed.

3) Encourage others, particularly the less active members, to offer their ideas. Remember, the group needs all the information it can get.

When your group reaches the point where each person can say, "Well even though it may not be exactly what I want, at least I can live with the decision and support it", then the group has reached consensus. This doesn't mean that all of the group must completely agree, but rather that everyone is in fundamental agreement. Therefore, treat differences of opinion as a way of 1) gathering additional information, 2) clarifying issues, 3) forcing the group to seek better information.

K. Additional Instruction for Leadership Conditions

For your task, you will have a leader, selected on the basis of your earlier ratings of one another's leadership abilities. The leader for the discussion is

NAME AND NUMBER OF SELECTED LEADER, BASED ON RANKING BEFORE BREAK, PRINTED HERE

Your leader has certain responsibilites and authority:

1. To decide the topics/items on which the group should focus its discussions at a particular time.

2. To summarize the group's progress or position from time to time.

3. To request members to move items in their lists when agreements have been reached.

L. Final Ranking Instructions

We are going to ask you to rerank the items now that you have had the discussion.

1. First, we will ask you to type in YOUR BEST ESTIMATE OF THE DECISION OF THE GROUP AS A WHOLE about the rank order of the items. Remember, use the ranks from 1 the most important, to 15 the least important, for the relative importance of each item for the survival of your group, ACCORDING TO YOUR PERCEPTION OF WHAT THE GROUP DECIDED.

2. Then, you will be asked to type in the order which is YOUR OWN FINAL DECISION ON THE RANK ORDER OF THE ITEMS. Remember, use the ranks from 1 the most important, to 15 the least important, for the relative importance of each item for the survival of your group, ACCORDING TO WHAT YOU, YOURSELF, REALLY THINK THE PROPER RANKING OF THE ITEMS SHOULD BE, now that you have had the discussion.

We suggest that you pencil in your rankings on the list below, before typing in the orders.

QUESTIONNAIRE FOR GROUP DISCUSSION PARTICIPANTS

NAME/# _____

DATE _____

CONDITION: _____

GROUP: _____

Please answer all of the following questions as honestly and carefully as you can.

The first three questions relate to the problem, and should be answered on the basis of your reactions as you read through it. These questions contain a number of rating scales on which you are to indicate your impressions of the problem by circling the number which best represents your answer.

1. The problem was:

	(26)	(56)	(27)	(5)	(2)	(4)	(0)	
:	1	2	3	4	5	6	7	: X=2.3
	Completely Interesting			Neutral			Completely Boring	

2. The situation struck me as:

	19	34	24	19	14	6	4	
:	1	2	3	4	5	6	7	: X=3.1
	Realistic						Unrealistic	

3. The issues involved were:

	27	40	29	11	11	2	0	
:	1	2	3	4	5	6	7	: X=2.5
	Completely Clear						Completely Unclear	

The next questions ask you to think about the group discussion system used today and to rate it on a one to seven scale for how satisfactory it would be for each of the following kinds of activities or processes. For each question a rating of 1 means Completely Satisfactory; a rating of 4 is Neutral and a rating of 7 would be Completely Unsatisfactory.

	Completely Satisfactory						Completely Unsatisfactory	
4. Giving or receiving information	17 : 1	45 : 2	26 : 3	13 : 4	14 : 5	5 : 6	0 : 7	2.8
5. Problem solving	2 : 1	30 : 2	23 : 3	18 : 4	26 : 5	18 : 6	2 : 7	2.8
6. Bargaining	4 : 1	26 : 2	27 : 3	13 : 4	31 : 5	12 : 6	6 : 7	3.8
7. Generating ideas	20 : 1	40 : 2	15 : 3	21 : 4	12 : 5	10 : 6	1 : 7	3.0
8. Persuasion	2 : 1	21 : 2	26 : 3	20 : 4	30 : 5	16 : 6	5 : 7	4.0
9. Resolving disagreements	1 : 1	17 : 2	26 : 3	26 : 4	30 : 5	17 : 6	3 : 7	4.1
10. Getting to know someone	4 : 1	14 : 2	20 : 3	26 : 4	27 : 5	15 : 6	14 : 7	4.3
11. Giving or receiving orders	26 : 1	30 : 2	25 : 3	12 : 4	18 : 5	7 : 6	2 : 7	3.0
12. Exchanging opinions	23 : 1	42 : 2	21 : 3	18 : 4	9 : 5	7 : 6	0 : 7	2.7

The following questions deal with your feelings about your group and its discussions and your participation today.

Once again, we ask you for a rating of between 1 (top rating) and 7 (bottom rating)

13. Taking part in this research was:

59	36	13	4	7	1	0	
: 1 :	2 :	3 :	4 :	5 :	6 :	7 :	1.9
Pleasant			Neutral			Unpleasant	

14. How satisfied are you with your own performance in this group discussion?

17	43	29	19	11	1	0	
: 1 :	2 :	3 :	4 :	5 :	6 :	7 :	2.7
Completely Satisfied						Completely Unsatisfied	

15. Did your group reach a consensus?

3	36	21	8	11	9	1	
: 1 :	2 :	3 :	4 :	5 :	6 :	7 :	2.6
Definitely Yes						Not at all	

16. Do you agree or disagree with the decision arrived at by the group?

20	46	27	10	11	4	1	
: 1 :	2 :	3 :	4 :	5 :	6 :	7 :	2.7
Strongly Agree						Strongly Disagree	

17. The general feeling of our group was:

	1	2	3	4	5	6	7	
	26	40	12	5	2	0	0	1.7
	Friendly						Unfriendly	

18.

	1	2	3	4	5	6	7	
	49	37	29	4	0	1	0	1.9
	Interested						Uninterested	

19.

	1	2	3	4	5	6	7	
	33	36	27	20	3	1	0	2.4
	Productive						Unproductive	

Finally, we need some background information.

20. Your age

(1) 3 Under 25

(2) 53 25 - 34

(3) 42 35 - 44

(4) 17 45 - 54

(5) 5 55 - 64

(6) ____ 65 & over

21. Your sex

(1) 82 Male

(2) 38 Female

22. Your highest educational level

(1) 0 Less than High School

(2) 2 High School graduate

(3) 13 Some College

(4) 38 4 Year College Grad.

(5) 44 Master's Degree

(6) 23 Doctorate

23. How well do you type?

(1) 29 Hunt and Peck

(2) 44 Rough or casual typing

(3) 30 Good typing (30 wpm, error free)

(4) 17 Excellent typing

24. How frequently have you used computer terminals in the past, for any kind of application?

(1) 15 Never

(2) 15 Once or twice

(3) 18 Three - Ten times

(4) 72 Frequently

REFERENCES

Bernard, H. Russell, Peter D. Killworth, and Lee Sailor
1979 An Experiment on the Degradation of Accuracy in Human Recall of Communications. Technical Report BK-118-79, Office of Naval Research, Code 452.

von Bertalanffy, L.
1968 General System Theory. New York: Braziller.

Borgatta, Edgar F. and Bales, Robert F.
1953 "Some findings relevant to the great man theory of leadership", American Sociological Review, 19 (1954):755-759.

Carey, John
1980 "Paralanguage in computer-mediated communication". Proceedings of the 18th Annual Meeting of the Association for Computational Linguistics and Parasession on Topics in Interactive Discourse, Philadelphia, Pa., June 19-22:67-69.

Dalkey, Norman C.
1969 The Delphi Method: An Experimental Study of Group Opinion. Santa Monica, Cal.: The RAND Corp., RM-5888-PR.

Dalkey, N.C., Brown, B., and Cochran, S.
1970 The Delphi Method, IV: Effect of Percentile Feedback and Feed-In of Relevant Facts. Santa Monica, Cal: The RAND Corp., RM-6118-PR.

Dickson, Gary R., Senn, James A. & Chervany, Norman L.
1977 "Research in management information systems: The Minnesota experiments". Management Science, 23,9:913-923.

Eady, Patrick Morton and J. Clayton Lafferty
1975 "The Subarctic Survival Situation". Plymouth, Michigan: Experiential Learning Methods.

Finn, Thomas Andrew
1982 Process and Structure in Computer-Mediated Communication. Dissertation proposal, Dept. of Psychology, Washington University.

French, John R.P. Jr.
1941 "The disruption and cohesion of groups," Journal of Abnormal and Social Psychology, 36:361-377.

Hare, A. Paul

1976 Handbook of Small Group Research, Second edition. New York: The Free Press.

Hiltz, Starr Roxanne, Kenneth Johnson, Charles Aronovitch and Murray Turoff
1980 Face to Face Vs. Computerized Conferences: A Controlled Experiment. Newark, NJ: Computerized Conferencing and Communications Center, NJIT: Research Report No. 12.

Hiltz, Starr Roxanne and Murray Turoff
1978 The Network Nation: Human Communication via Computer. Reading, Mass.: Addison Wesley Advanced Book Program.

Johansen, Robert, Jacques Vallee and Kathleen Spangler
1979 Electronic Meetings: Technical Alternatives and Social Choices. Reading, Mass., Addison Wesley Publishing Co.

Hiltz, Starr Roxanne, Turoff, Murray, and Johnson, Kenneth
1983 Mode of Communication and the "Risky Shift": Controlled Experiments with Computerized Conferencing and Anonymity in a Large Corporation. Forthcoming research report.

Johnson-Lenz, Peter and Trudy
1981 "Consider the groupware: Design and group process impacts on communication in the electronic medium". In Hiltz, Starr Roxanne and Kerr, Elaine B., eds., Studies of Computer Mediated Communication Systems: A Synthesis of Findings. Final Report to the National Science Foundation.

Keen, Peter G.W. and Morton, S
1978 Decision Support Systems. Reading, Mass.: Addison Wesley.

Kelly, C.W., Chase, L.J., and Tucker, R.K.
1979 "Replication in Experimental Communication Research: An Analysis". Human Communication Research, 5,4:338-342.

Kerr, Elaine B. and Hiltz, Starr Roxanne
1982 Computer-Mediated Communication Systems: Status and Evaluation. New York: Academic Press.

Linstone, Harold A. and Turoff, Murray
1975 The Delphi Method: Techniques and Applications. Reading, Mass.: Addison-Wesley Advanced Book Program.

Lipinski, Hubert, Spang, Sarah, and Tydeman, John.
1980 "Supporting task focussed communication", in Benenfeld, Alan R., and Edward John Kazlauskas, eds.,

Communicating Information: Proceedings of the 43rd ASIS Annual Meeting. New York: Knowledge Industry Publications, Inc.

Maier, Norman R.F. and Allen R. Solem
1952 "The contribution of a discussion leader to the quality of group thinking: The effective use of minority opinions." Human Relations, 5:277-288.

Mowshowitz, Abbe
1976 The Conquest of Will: Information Processing in Human Affairs. Reading, Mass.: Addison Wesley.

Nixon, H.L.
1979 The Small Group. Englewood Cliffs, N.J.: Prentice Hall.

Palazzolo, Charles S.
1981 Small Groups: An Introduction. New York: D. Van Nostrand Co.

Scher, Julian M.
1981 "Distributed decision support systems for management & organizations", in Donovan Young and Peter G.W. Keen, eds., DSS-81 Transactions. Execucom Systems Corporation: 130-140.

Short, John, Williams, Ederyn, and Christie, Bruce.
1976 The Social Psychology of Telecommunication. New York: John Wiley and Sons.

Turoff, Murray
1974 "Computerized conferencing and real time delphis: unique communication forms". Proc. 2nd International Conference on Computer Communications: 135-142.

Uhlig, Ronald P., Farber, David J., and Bair, James H.
1979 The Office of the Future. Amsterdam: North Holland.