COMPUTERIZED CONFERENCING & COMMUNICATIONS CENTER

at

NEW JERSEY INSTITUTE OF TECHNOLOGY

FACE-TO-FACE VS. COMPUTERIZED CONFERENCES:
A CONTROLLED EXPERIMENT

BY:

STARR ROXANNE HILTZ, KENNETH JOHNSON, CHARLES ARONOVITCH
and MURRAY TUROFF

VOLUME I: FINDINGS

AUGUST 1980

RESEARCH REPORT NUMBER 12

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

FACE TO FACE VS. COMPUTERIZED CONFERENCES; A CONTROLLED EXPERIMENT

Starr Roxanne Hiltz, Kenneth Johnson, Charles Aronovitch

and Murray Turoff

FINAL REPORT

August 1980

The research reported here is supported by a grant from the Division of Mathematical and Computer Sciences (MCS 78-00519). The findings and opinions reported are solely those of the authors, and do not necessarily represent those of the National Science Foundation.

We are grateful to Julian Scher and Peter and Trudy Johnson-Lenz for their contributions to the design of the experiments; to John Howell and James Whitescarver for their programming support; and to our research assistants for their dedicated efforts in carrying out the experiments and coding questionnaires: Joanne Garofalo, Keith Anderson, Christine Naegle, Ned O'Donnell, Dorothy Preston, Nancy Rabke, Karen Winters, and Stacy Simon. The Kendall's Coefficient analysis was carried out by Christine Naegle; the analysis of the relationship between social-emotional positive Bales units during training discussions and subsequent degree of consensus by Keith Anderson; and the analysis of sex and opinion change by Ned O'Donnell. Nancy Rabke aided greatly in the work of preparing the final report. Marion Whitescarver has our appreciation for her assistance in keying in and editing this report.

We would also like to thank Robert Bales and Experiential Learning Methods for their cooperation in providing documentation and permission to use adaptations of problem solving tasks which they originally developed.

TABLE OF CONTENTS

LIST OF TABLES

FIGURE PAGE

ABSTRACT

A controlled laboratory experiment compares face to face with computerized conferences for two different types of group decision tasks. One problem is a structured, rank ordering task which requires knowledge pooling. The other problem is an unstructured, value laden human relations task.

Various measures of the process and outcome of group decision making were measured for the sixteen groups of five members each. Among the significant findings are that

.There is no difference in the quality of solution reached between the two modes of communication.

.Face to face groups are significantly more liklely to be able to reach total consensus on the solution to a problem.

.Dominant individuals are more likely to arise in face to face groups.

.There are two to three times as many communication units generated in face to face meetings as in computerized conferences, within the same time period.

.There are significant differences in Interaction Profiles between the modes of communication. These differences are correlated with

differences in the quality of solution and consensus outcomes.

.New users of computerized conferences find face to face conferences more satisfactory for most communications tasks, but tend to rate computerized conferences on the satisfactory side of neutral.

Many of the above findings are influenced strongly by task.

CHAPTER ONE

INTRODUCTION

This is a report on the first controlled experiment conducted as part of a four year effort to explore the feasibility and effectiveness of using the computer to structure the communication for various types of group tasks. It uses a language called INTERACT, developed as part of this grant effort, to administer all instructions and conduct a group problem solving discussion in a computer conference. The experiment compares the process and outcome of face to face vs. computer mediated group problem solving discussions.

The objectives of this experiment are the following "basic research" questions:

1) To gain quantified and detailed knowledge about the consequences and characteristics of computerized conferencing as a communications mode, as compared to the usual face-to-face discussion mode.

2) To lay the foundation for a subsequent experiment which will seek to alter the process of group communication via computer, in order to improve group performance.

3) To assess the feasibility of using a high level language to conduct automated experiments on group communication and problem solving.

The experiment uses a human relations problem developed by Robert Bales and a complex group ranking problem. To code process of interaction, it uses the classic Bales Interaction Process Analysis technique. We took advantage of available documentation and results on other Bales type experiments and the results of a pilot study sponsored by the Division of Mathematical and Computer Research to provide the experimental procedures and some comparison data. (See Hiltz, 1975 and Hiltz, Johnson and Agle, 1978, for a summmary of the earlier work and pilot study which formed the basis for the design and objectives of this study.)

Brief Description of EIES

The host for these experiments is EIES, the Electronic Information Exchange System, built and operated at the New Jersey Institute of Technology with the support of grants from the National Science Foundation. EIES is primarily a communication medium. It allows over 500 scientists who are geographically dispersed throughout North America and in several other nations to communicate with one another on a continuing basis. The EIES users are organized into groups which share common interests and tasks. They can communicate by typing into and reading from a computer terminal, using messages, group conferences for seminar-like discussions, and notebooks for remote co-authoring. There are many other systems which incorporate some of these features (See Hiltz and Turoff, 1978, for a complete description of EIES and similar systems). However, EIES is also

designed as a "laboratory without walls" for the study of communication. A programming language named "INTERACT" can be used for such purposes as altering the interface between user and computer, collecting data on the communications which occur, or designing and administering questionnaires or experimental procedures. INTERACT was used in this experiment to create a simple four-command interface for the subjects, to administer all instructions, and to isolate them from other activities on the EIES system. The subjects were not involved with any of the people or activities on the system; they were concerned only with their own group and its conference.

Overview of the Experiment

The chief independent variable of interest is the impact of computerized conferencing as a communications mode upon the process and outcome of group decision making, as compared to face-to-face discussions.

In computerized conferencing, each participant is physically alone with a computer terminal attached to a telephone. In order to communicate, he or she types entries into the terminal and reads entries sent by the other participants, rather than speaking and listening. Entering input and reading output may be done totally at the pace and time chosen by each individual. Conceivably, for instance, all group members could be entering comments simultaneously. Receipt of messages from others is at the terminal print speed of 30 characters per second.

Even though all five participants were on-line at the same time, there is considerable lag in a computer conference between the time a discussant types in a comment, and when a response to that comment is received. First, each of the other participants must finish what they are typing at the time; then they read the waiting item; then they may type in a response; then the author of the original comment must finish his or her typing of a subsequent item and print and read the response. There is thus a definite "asynchronous" quality even in "synchronous" computer conferences. As a result, computer conferences often develop several simultaneous threads of discussion that are being discussed concurrently, whereas face to face discussions tend to focus on one single topic at a time and then move on to subsequent topics. A variable of secondary interest is problem type. Much experimental literature indicates that the nature of the problem has a great deal to do with group performance. One type of problem that we used is the human relations case as developed by Bales. These are medium complex, unsettled problems that have no specific "correct" answer. The second type was a "scientific" ranking problem (requiring no specific expertise), which has a single correct solution plus measureable degrees of how nearly correct a group's answer may be. The ranking problem, "Lost in the Arctic", was adapted for administration over a conferencing system by permission of its originators (See Eady and Lafferty). After rejecting three other ranking problems in pretests, we found that Arctic satisfied all five of our criteria: 1) It was interesting; 2) doable in 90 minutes or less; 3) possessed a criterion; 4) produced variation in the quality of solution reached by test groups; and 5) subjects were unlikely to have previously encountered it.

The experiments thus had a basically 2 x 2 factorial design (see figure one). The design and the analysis are explained further in an appendix to this chapter. The factors were mode of communication (face-to-face vs. computerized conference) and problem type (human relations vs. a more "scientific" ranking problem with a correct answer). These factors constituted the "independent variables". The group size was five.

In order to decrease subject variability and fatigue, subjects were trained for one week before the experiment and administered a "test" of their ability to enter and read comments on the system. Six to seven subjects were trained, and five were selected. Besides minimal competency levels, an additional selection criterion was a desire to have at least one male and one non-white subject in each group of five. The subjects were Upsala College students, including many continuing education students who were older than "normal" college age. During the second week, each group was run through one problem in one mode, given a short coke and cookie break, run through the second problem in the second mode, administered post-test questionnaires, and debriefed. The experiments were carried out during the summer and fall of 1978.

Within each block, each group was randomly assigned to one of the four possible combinations of order of problem and order of mode. The experiment took about four hours to run, and involved a large number of instructions and actions by the experimenters and assistants, conditional upon the sequence to which the group was

qssigned. The procedures are described more fully in a subsequent chapter on methodology, and complete details are included in the Appendix.

Figure 1-1

Design of Experiment One

Two By Two Factorial with Repeated Measures: Blocks of Four

	Task Type A	Task Type B
	Groups	
Face-to-Face	4	4
Computerized Conference	4	4

Notes: Each group had two tasks in two different modes. In each block of four groups, groups were randomly assigned to begin in one of the four conditions; then they did the other problem in the other mode. Thus, all conditions had a total of eight groups. Group size was five.

Dependent Variables

The dependent variables we are focusing on are:

1. Quality of Decision

2. Ability to Reach Consensus

3. Subjective satisfaction with the communication media

The aspects of the communication process are conceptualized as intervening variables:

1. Amount and type of communications which we coded using Bales Interaction Process Analysis (see Figure 2).

2. Inequality of participation or dominance by a single "leader".

We also have a number of covariates, including sex of participants.

In the remainder of this introductory chapter, we will briefly review the literature that led to selection of the variables, and list the hypotheses with which we started. The project began with a complete review of all literature on small group problem solving which might be relevant to controlled experiments focussing on the effects of CC as a mode of communication (see Hiltz,1975). From this literature review, a small set of variables and measures was isolated which

appeared most promising for this initial experiment. The sections which follow summarize that part of the literature which led to the development of our hypotheses and procedures.

Figure 2

Categories in Bales Interaction Process Analysis

1. Shows solidarity
2. Shows tension release, jokes
3. Agrees
4. Gives Suggestions
5. Gives Opinions
6. Gives Orientation
7. Asks orientation
8. Asks opinion
9. Asks suggestion
10. Disagrees
11. Shows tension
12. Shows antagonism

Source: Bales, 1950

The Selection of Problem Types

A widely used classification of task types was presented by Shaw (1963), who identified ten potential task dimensions through a review of the literature. Judges used an adaptation of a Thurstone scale (a ranking technique) to sort 104 tasks along these ten dimensions. What emerged were three factors, when a factor analysis was performed: Task difficulty, solution multiplicity and cooperation requirements.

Difficulty was defined as the amount of effort required to complete the task, as determined by such dimensions as the number of operations, skills and knowledge required.

Solution multiplicity was defined as the degree to which there is more than one correct solution to the task. It is a complex dimension involving number of alternatives for task completion, and the degree to which acceptable solutions can be verified.

Cooperation requirements were defined as the degree to which integrated interaction of group members is required to complete the task. Tasks which do not require group cooperation could be completed by each group member working independently and at his own speed.

It was our desire to find two task types which are both complex and require cooperation, but which differ on solution multiplicity and

verifiability. Within this "difference", we wanted one set of tasks to involve a ranking type operation which would be amenable to exploration in later experiments with augmentation of group problem-solving using a computerized decision aid. Secondly, we wished one task to seem to be a "human relations" type, and the second to seem more scientific or technical.

Based on our own pilot studies and previous experiments, we settled upon a Bales human relations task ("Forest Ranger") and Hall's "Man on the Moon" task as two problems which are both complex, and which both involve instructions that the group must cooperate because its task is to reach consensus. Pretests proved that "Lost on the Moon" had been seen by many potential subjects; "Arctic" was eventually selected as a ranking problem that met all the criteria outlined above.

Background: The Bales Experiments and Interaction Process Analysis

Working at the Laboratory of Social Relations at Harvard, Bales and his colleagues developed a set of categories and procedures for coding the interaction in small face-to-face decision-making groups which became very widely utilized and generated a great deal of data about the nature of communication and social processes within such groups.

Coding of the communications interaction by Interaction Process Analysis involves noting who makes a statement or non-verbal participation (such as nodding agreement); to whom the action was addressed; and into which of twelve categories the action best fits (see Figure 3).

Bales and his colleagues have established that for small groups asked to discuss a complex human relations problem with no clear "solution" or "answer", there emerges both a fairly standard distribution of types of contributions and also clear "phase" movements and regularities.

Interaction Process vs. Outcome

As Hackman and Morris (1975) state in their review, "research on group effectiveness rarely includes explicit quantitative assessment of how group interaction affects group performance"(p.3). For the

ranking task, we will have outcome measures for quality of solution and degree of consensus reached. We will also have interaction process measures in the form of percentage distributions for the Bales categories. Thus, we will be able to examine not only how medium affects process, but also how these differences in process in turn affect the outcome of the group decision making.

The few studies that have been done lead to the prediction that we will find significant process-outcome relationships. For example, Katzell, Miller, Rotter and Venet (1970) used Interaction Process analysis in a "20 questions" type of task, and found some process-performance relationships. For example, as seeking information and giving information increased, time to solution increased. Hackman and Morris summarize some very strong correlations obtained between a sixteen category coding scheme and eight outcome criteria (Hackman and Morris, 1975, pp 9-11). The interaction coding is similar to the task oriented categories in Bales IPA, but more finely detailed. For example, "clarify" and "repeat" are separate categories. The dependent or criterion variables include dimensions such as length, originality, and adequacy of the solutions.

The development of Interaction Profiles for the computer conferencing condition will enable us to quantify just how the content and sequence of group communications differ in the computer conference communications mode as compared to the face-to-face conference. There have been subsequent modifications to Bales IPA, but we decided to stay with the well documented and widely used original version (Bales, 1951, available in paperback). There are some predictions in

previous work about what kinds of differences could be expected to occur. For example, Vallee et. al. (1974, p.92) said that they observed more questions asked in face-to-face meetings than in FORUM computer conferences. However, this did not conform to our casual observations. It was decided to make the predictions of significant differences in interaction profiles non directional.

Inequality of Participation

One standard mode of assessment of group interaction utilized by Bales and his colleagues is the "who-to-whom" matrix, with the originators of statements designating a series of rows, and the recipients, the columns.

It was found that if the...

> Participants are ranked by the total number of acts they initiate, they will also tend to be ranked: a) by the number of acts they receive; b) by the number of acts they address to specific other individuals; and c) by the number of acts they address to the group as a whole. (Bales, et al., 1951, p. 468.)

There usually emerges a "top man" who sends and receives a disproportionate number of messages, and who...

> a) addresses considerably more remarks to the group as a whole b) receives more from particular others than he gives out to them specifically (Bales, et al., 1951, p. 465)

> Commenting on the processes which produce this dominance, Bales (1955, p. 34) has written:
>
> This tendency toward inequality of participation over the short run has cumulative side effects on

> the social organization of the group. The man who gets his speech in first begins to build a reputation. Success in obtaining acceptance of problem-solving attempts seems to lead the successful person to do more of the same, with the result that eventually the members come to assume a rank order by task ability. In some groups, the members reach a high degree of consensus on their ranking of "who had the best ideas". (The members are interviewed by questionnaire after each meeting.) Usually, the persons so ranked also did the most talking and had higher than average rates of giving suggestions and opinions.

Other experiments have also found that the amount and type of communicating which a person does in a face-to-face group discussion involving problem-solving is strongly related to the probability of being perceived as a "leader". Some studies and coefficients of correlation obtained include:

> a) Norfleet (1949), using Bales IPA, found correlations of .94 and .95 between relative rank on amount of participation (communication) and relative rank on perceived productivity among group members.
> b) French (1950) found a correlation of .96 between time spent talking and ratings of leadership.

Experience during the pilot studies and theories and findings in more recent work that follows up on Bales' studies indicated that level of participation should be conceptualized and analyzed in terms of three dimensions (Burke, 1974, 832-833):

1) The number of times that an individual participates, or the number of "turns".

2) The amount of participation on each turn, which can be measured by Bales interaction units, or by length of time spent speaking or

number of lines or words composed in a written form of communication.

3) "Back-channel" or non-verbal participation, which often functions to give turn-yielding or turn-suppressing signals.

The first two aspects of participation can be most easily quantified and used as dependent variables. In the computerized conferencing condition, the number of separate messages or conference comments corresponds to the number of turns. In the face-to-face condition, the number of turns can be coded from tape recordings.

Amount of participation can be measured by the number of Bales units coded as "from" each individual, in order to achieve the most comparable measure between the two medias.

"Back-channel", non-verbal communication was not coded, since there is no comparable information channel in the computerized conferencing condition.

Latency of Verbal Response, Dominance and Quality of Decision

What, then, causes a person to do most of the talking? The tendency for an individual to be slow in responding or jumping into a conversation, or prone to speedy replies and interruptions, was noted by Chappel and or Arensberg in 1940, and has come to be recognized as a fairly stable individual characteristic. It is called "latency of verbal response" (L.V.R.), and is measured by response time on sentence stub completion tasks. For example, in a task which

minimized differences in competence (moral dilemmas, such as whether a man with a wife dying of cancer should steal some expensive drug which might save her), Willard and Strotbeck (1972) found that a participant's L.V.R. was the strongest predictor of participation (correlation of -.60, compared with measure of I.Q. and personality, while the correlation between I.Q. and percent participation, for instance, was only .12).

What is interesting here is that the evidence indicates that persons who happen to be "fast on the draw" in a face-to-face verbal situation, and who may not be particularly intelligent or correct, tend to dominate the discussion and decision-making process in small groups. Computer conferencing as a mode of communication would pretty much suppress L.V.R. as an operative variable, it is hypothesized, and the relative verbosity of a person in written communication is much more likely to be resented than unconsciously deferred to. Thus, it is quite possible that intelligence and correctness might be much more highly correlated with the leadership and dominance processes in decision-making that developed in a computer conferencing group.

The Functions of Inequality

Burke offers a theoretical explanation of the interrelation among various factors found to be associated with inequality of participation, and argues for its functional necessity in enabling a group to reach consensus, as follows (Burke, 1974, 842-843):

> Achieving coordination and consensus requires a manner of participation, which results in inequality. Whoever plays the coordinator's role probably does so by involving himself in interchanges with others to solicit, respond to, offer, and integrate ideas and opinions on the topic at hand, to the extent that: (1) a group member does this, (2) the interchanges are . . . organized such that the floor is usually returned to him, and (3) he initiates these interchanges because his low verbal latency enables him to grab the floor (then): (1) he will be active, (2) most of his turns will be given to him (rather than "stolen"), (3) he will be perceived as the leader . . . and (4) he will have a low verbal latency.

Burke further argues that the inequality of participation which characterizes this process is necessary in order for the group to become organized enough to reach a consensus on how to solve a problem.

Many persons who have not observed group decision-making processes conducted in other than face-to-face discussions tend to think that it will be difficult or impossible for members to understand and interact with one another without the various cues provided by such "back-channel" communication as facial expression. However, the existing experimental evidence indicates that this is not the case, and that indeed, most problem-solving can be done as well or better in non-face-to-face conditions. For example, Williams(1975) found that mode of communication (face-to-face vs. audio-only conference vs. closed circuit TV) had no effect on either number of ideas generated or originality and quality of ideas generated (as judged by raters). Werner and Latane (1976) compared face-to-face, TV, audio, and handwritten conditions. They found that "The communications

medium used for discussion tended to be less important than a partner's responsiveness in determining opinion changes and reactions to discussions. The media did not differ in their ability to convey positive images or to impart pleasure to the interaction."

We thus arrived at the predictions, based on the literature and previous pilot studies, that computerized conferencing, as compared to face-to-face discussions, will probably result in more equal participation, and that this, in turn, is likely to lead to the generation of more ideas and suggestions on how to solve a problem, but less likely to lead to reaching total consensus on a decision in a given amount of time, since it is less likely that a single leader will emerge to push the group towards agreement. A related factor is that the absence of non-verbal communications makes it much easier for a "deviant" group member to hold out against the other members of the group, rather than go along with the group. In the pilot studies, there were no groups in the face-to-face condition in which a 4-1 split was maintained; the deviant always reluctantly went along. In the computerized conferencing condition, there were several instances of a stable, adamant 4-1 or 3-2 split, with the deviant steadfastly holding that he or she did not agree with the rest of the members. We thus predicted that the computerized conferencing condition would be characterized by a lower probability of reaching a total consensus as compared with face to face discussions.

Related Experiments in Telecommunications

This study has built upon some of the measures and concepts used by the Communications Studies Group (CSG) in Great Britain (See Short, Williams, and Christie, 1976, for a comprehensive and very readable summary of this work). Many of the experiments conducted by this group compared various modes of communication for various types of group tasks, as does the research reported here. CSG studies included face to face, audio conferencing, and video conferencing modes. We have used scales developed by CSG in measuring subjective satisfaction with the media.

The only other controlled experiment we are aware of which compares face to face and computerized conferencing modes of communication was carried out by T.N. Westgate at the Cranfield School of Management in Great Britain during 1977 (Westgate, 1978). The pilot series of experiments used 32 subjects engaged in a crisis negotiation exercise. Westgate borrowed some of the same CSG scales as are employed in this study to measure attitudes toward the media, so that some comparison of our findings to his will be possible.

Time and Medium

Though there had been no studies directly comparing face to face with computerized conferences at the time this study was designed, some previous work comparing communication modes was suggestive. Chapanis and his colleagues have compared face-to-face with audio and

slaved-typewriter written communications. They found that "communication by voice is much more rapid and wordy than is communication by typewriter" (Chapanis and Overby, 1974; Chapanis, 1975). Two slaved typewriters bear little resemblance to a computerized conference in which five persons conceivably might be typing at once (since in a slaved typewriter condition, the "recipient" must sit and receive communications one letter at a time as they are typed). However, pilot studies did indicate that, at least with neophytes, groups using computerized conferencing often seemed to need longer than the forty minutes alloted by Bales for face-to-face discussions. Therefore, we allowed 60 minutes for this problem and 90 minutes for the more difficult "Arctic" problem. Within this time frame, Chapanis' work and other previous pilot studies led to the prediction that there would be more communication units in the face-to-face condition.

Gender

This variable is of secondary interest in this study. However, its influence will be explored to the extent that is possible.

Theoretical investigations of the effects of "irrelevant" statuses upon expectations and behavior can be traced back to Hughes (1945), who described the hypothetical situation that might occur in interaction between a black female physician and a white male office-worker. He argued that such diffusely evaluated characteristics were important variables in determining "subordination and superordination" in interaction processes. Merton (1968) and others have also analyzed the impact of the degree of "salience" and "dominance" of various ascribed roles within the role

set. A very large number of experimental studies have substantiated the theoretical generalization that evaluated statuses have a very diffuse and strong effect upon dominance (or "leadership") in group interaction, regardless of whether or not these characteristics are relevant to the task at hand. For example, Strotbeck et. al.(1958) analyzed the effect of gender and occupation upon jury deliberations. In Torrence's (1954) study, bomber crews formed expectations about performance based on relative rank, regardless of how irrelevant such rank was to tasks such as dot estimation or "horse trading".

Some recent work by Berger and Webster and their various co-authors provides a plausible explanation of the process. (Berger et. al., 1956; 1972; 1976; Berger and Fisek, 1970; 1974; Webster, 1974). The findings are summarized as follows (Webster, 1977, p. 42): In small groups engaged in problem-solving interaction, the members come to reach conclusions about the relative problem-solving ability of each person called an "expectation state", which determines whose opinions they want to hear and the evaluation they will make of suggestions of uncertain quality.

> This formulation posits a two-step process to explain the effects of status characteristics upon interaction: (1) actors notice the various characteristics each member possesses and on the basis of these characteristics form performance expectations for the task at hand; (2) these expectation states, once formed, produce the observable inequality of behavior between members.

We are conceptualizing gender as just such a task-irrelevant status,

which, in the face-to-face condition, is likely to strongly affect the amount and type of participation.

A wide variety of studies in the sex-role literature show that in face-to-face mixed sex groups, females tend to participate less than males; in other words, males assume the leadership, or dominance roles and the females "conform". (See, for example, Nord, 1969, and Carpage and Lindskold, 1973). Moreover, contributions by females tend to be "devalued", that is, considered less useful or of lower quality than those made by males. As a result, the ideas and potential contributions of females are not fully utilized (McKee and Sherrifas, 1957; Goldberg, 1968).

In addition to amount of participation, we may expect to find some differences in the type of participation. Growing out of the Parsons and Bales traditions is the theory that there will be sex-typed participation roles (see Bales, 1949; Parsons et. al., 1953; Slater, 1955). As Meeker and Weitzel-o'Neill summarize the theory in their recent review (1977, p. 91):

> According to the general theory, task behavior (which is primarily in the attempted answer categories of the Bales coding system) and positive social behavior (primarily in the positive reactions categories of showing solidarity, tension release and agreement) are incompatible, but both are essential to a viable small group. A pattern of role differentiation, in which a group has one "task leader" with higher rates of task behavior than other group members, and a different "social leader", who has higher rates of positive social behavior than other group members, was hypothesized to be a universal feature of a viable small group. The sex role differentiation hypothesis was derived from this more general role differentiation hypothesis.

The tendency, according to theory, is for males to be the task-oriented leaders and women to be the "social leaders", because of sex role socialization. Some studies using IPA coding (such as Strotbeck and Mann, 1956) have supported this.

We follow Meeker and Weitzel-O'Neill's argument that, insofar as such processes occur, it is probably because of the following (ibid., p. 96):

> A task contribution by one member of a task-oriented group which is accepted by others will be assumed by both self and others to raise the status of the contributor. Raising one's own status is legitimate for persons with high external status, but not for those with low external status. Since men have higher status than women, raising one's own status relative to the status of others within a small group is legitimate for the former, but not for the latter.

We hypothesize that the "illegitimacy" or "social disapproval" for a low-status person seeking to take task-oriented leadership will not be adequately transmitted in the computerized conferencing condition without non-verbal cues, and that, therefore, sex-typed behavior will be less predominant in the computerized conferencing condition.

Other recent theoretical and empirical work indicates that the total group composition must be taken into account when assessing the impact of a "minority" status upon participation and ranking processes. When the "minority" becomes the "majority" (such as in a group that is 4-1 female), then it can be expected that social pressures that normally operate are much changed, and that the

non-dominant stratum will feel more free to take an aggressive leadership role. For example, Kanter (1977) draws the following distinctions (p. 965):

> Proportions, that is, relative numbers of socially and culturally different people in a group, are seen as critical in shaping interaction dynamics, and group types are identified on the basis of varying proportional compositions. "Skewed" groups contain a large preponderance of one type (the numerical dominants) over another (the rare "tokens") . . . Three perceptual phenomena are associated with tokens: visibility (tokens capture a disproportionate awareness share), polarization (differences between tokens and dominants are exaggerated), and assimilation (tokens' attributes are distorted to fit pre-existing generalizations about their social type). Visibility generates performance pressures . . .

Eskilson and Wiley (1976) used three person groups coded by Bales IPA and engaged in a face-to-face problem-solving situation and found that the traditional sex-role stereotypes were confirmed. For instance, males designated as leaders did engage in more "instrumental, leader-like" behavior, and females engaged in more "affective" activity. However, they found that sex composition was an important contextual variable. "For example, females leading two males performed minimal amounts of leader behavior" (Ibid., 92-93).

Drawing from these studies, we planned to analyze the group composition context for the effect of sex upon amount and type of participation. We expected that "token" situations (four to one ratios) would have a strong effect in face-to-face situations; but in the absence of cues and non-verbal pressures in the computerized conferencing condition, sex composition would probably not have much effect.

We also planned to look for interaction between sex, sex composition, and the task type. "Scientific" tasks are thought to be "male" in our society, whereas human relations tasks are more "female". We expected that there would be some differences in the amount and type of participation associated with task and sex in the face-to-face conditions, but not in the computerized condition. (One example of a study on the relationship among sex, task type, and performance is Milton, 1959, "Sex Differences in Problem Solving as a Function of Role Appropriateness of the Problem Content".)

INITIAL HYPOTHESES

The hypotheses listed below were formulated before actually recruiting the subjects or conducting the pilot tests of the experimental procedures. In the chapters that follow, most of the main hypotheses are tested as originally planned. As the study unfolded, we did develop a few hypotheses by generalizing or finding patterns from unexpected observations. Wherever data or generalizations refer to hypotheses that were not stated before the study was conducted, this is noted. In addition, some of these hypotheses were not tested, because of insufficient data or because the analysis proved to be very time consuming, and it was decided to forego it in favor of proceeding on to the next experiment.

Hypothesis 1

Better decisions would be generated by groups using computerized conferencing than by face-to-face groups. The dependent variable is percentage improvement in quality of decision. Quality of decision is measured by deviation from the criterion on the Arctic problem. Since some groups start out with a better average solution than others, we actually want to look at improvement in quality of decision as a result of the discussion.

Hypothesis 2

Computerized Conferencing will be less likely to result in consensus. The dependent variable is whether or not the group reached a unanimous decision, for the Human Relations problem (tested by X2). For the ranking problem it will be measured by Kendall's coefficient of concordance (see Chapter 5).

Hypothesis 3

Computerized Conferencing will produce a different distribution of statement types than face-to-face groups. This analysis will be repeated for each of the twelve categories. It is predicted that more opinions (or options) will be put forth in CC than FtF. The other predictions of differences are non-directional. (See Chapter 4 for results).

Hypothesis 4

There will be more equality of participation in computerized conferencing. (See Chapter 5).

Hypothesis 5

There will be an interaction of task or problem type with communication mode. Not enough previous research has been done to predict the interaction of problem type with communication mode. The

following potential effects will be analyzed in terms of the differences between problems in relation to communication mode:

a) interaction profile

b) inequality of participation

c) quality of decisions

d) degree of consensus

The above were our major hypotheses for these experiments. We also had a number of secondary hypotheses relating to sex and sex composition.

Hypothesis 6

There will be a greater equality of female participation in computerized conferencing.

Hypothesis 7

There will be some differences in the association of sex with IPA distributions. Specifically, females will be more likely to express disagreement in computerized conferencing than in face-to-face discussions (not yet tested).

Hypothesis 8

Sex composition of the group will interact with mode of communication to affect equality of participation by sex. It is when a female or male is in the "token" position that computerized conferencing will most affect participation (not tested; insufficient male subjects).

We had hoped to test the effects of the sex and race of individuals and sex and race composition of groups by purposely varying the composition of the subject groups. However, we had a very difficult time recruiting 80 subjects for this study, after rejecting those with insufficient typing skills, from the continuing education and regular enrollment students at Upsala. We did have at least one male in every group, but it was more difficult to recruit males than females, so we were not able to pursue the sex composition hypothesis. It was also difficult to recruit and train sufficient numbers of minority subjects (even though Upsala has about 20% minority enrollment), so that there is an insufficient number of minority subjects to test hypotheses about race.

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 first independent variable (A) is Mode of Communication; the second independent variable (B) is Order of the Problem (first or second) ;AxB means interaction between mode and order of problem. The problem itself, Arctic or Forest Ranger, is controlled by performing separate analyses for each problem. Factor C is "Group". The "within groups" (WG) or error variances shown in this report are the WG, C/AB, and S/ABC terms. The error variance is due to factors other than the treatment conditions.

Data such as that obtained in this experiment are generally analyzed with the analysis of variance techniques. However, there may be some legitimate question as to whether some of the data meets the assumption of interval level of measurement necessary to perform the analysis of variance. In order to be sure that obtained differences were due to treatment effects and not violation of the assumption of interval level of measurement, all significant treatment effects were also analyzed with appropriate nonparametric analyses which require only ordinal level of measurement. Where the analyses resulted in different results, both are reported.

The basic design for the experiment was a factorial design with interactions that were partially confounded. Normally in analysis of variance designs each subject serves in only one treatment condition (completely random design) or in all treatment conditions (correlated design). A commonly used design which combines the above designs within a single design is the mixed factorial design in which one variable is completely random in nature and the other is correlated.

If we had randomly assigned our subjects to either the face to face condition or the computer conference condition and then repeated measures over just the problem condition, then this would have been a mixed factorial design. Instead, we chose to repeat measures over both variables. Thus, each group would be exposed to both modes of communication as well as both problems. If one group received the Arctic problem in the face to face condition, then they would receive the Bales problem in the computer conference condition. The other group in this set would then receive the problems under the opposite communication conditions. Two other groups would then receive the same treatment conditions, but in reversed order. Differences between the groups in this design will form a part of the interaction. Thus interaction components will be confounded by the group effects. This design is generally reserved for situations where some information about interactions is sacrificed in order to gain greater power for interpreting the noninteraction components with a given number of subjects.

However, in this experiment this was not the primary reason for selection of such a design. The subjects were trained in the use of the computer terminal a week before they took part in the problem discussion. To have trained only the CC groups would have severely confounded "training" or "practice" effects with treatment effects. Since all subjects were trained to use the terminals, they undoubtedly expected that they would use the terminals in the experiment. Thus, in order to insure subject satisfaction, subjects were all given two problems to solve, one in each of the treatment conditions.

In the primary analyses we do not actually compare the different problems in the same analysis of variance, and thus never really analyze the confounded design. Instead, we look at the FtF vs CC within each problem type. Thus we are left with a design in which mode of communication is the major variable. We cannot ignore the fact that, for some subjects, the problem they are solving is their first problem, while for others, it is the second. Thus problem order becomes a second variable.

One might analyze the data with a 2 x 2 factorial design in which the group is the unit of analysis. This would be a legitimate design, but one that is not overly powerful. In the 2 x 2 factorial design one has one df (degree of freedom) for each of the treatment conditions (as with each of these designs) and 12 degrees of freedom for the error (WG) term, for a total of 15 degrees of freedom (or N-1 df).

Even though that design is statistically correct, it ignores the fact that there are 5 subjects in each group. This is called a nested design, because the effects that occur in each group are unique and nested under both the mode of communication and problem order variables. The actual design then is a 2 x 2 x 4 nested factorial design where the first factor is mode of communication (A), the second variable is problem order (B), and the third variable (Group, or "C") is the nested one (C/AB). The nested design has no inherent advantage over the 2 x 2 design. It simply separates out another source of variation and it allows one to see if there are indeed

different (unique) things occurring within each group. If there are nested effects, then the design becomes the equivalant of the 2 x 2 factorial. In the nested design the error term for the first two factors and the interaction is the C/AB (nested) term. Thus there are one and 12 degrees of freedom (df) for the F tests for the A, B, and A x B effects, as in the 2 x 2 factorial design. The error term for the nested effect (C/AB) is the S/ABC term. Thus there are 12 and 64 df for the nested term.

Given the above description of the nested design, it may not be obvious why one would choose to pull out a source of variation which is not of any particular interest. However, if the nested effect is not significant, then we can pool the two sources of error (the C/AB and S/ABC) in the design and obtain a total of 76 degrees of freedom for the error term, or the equivalent of having a total of 80 independent observations, instead of the 16 observations analyzed in the 2 x 2 factorial design. This new pooled error term is then used for the analysis of the A, B, and A x B effects. Thus the pooled design derived from the nested design has considerably greater power than the 2 x 2 factorial design. In the results, both designs are reported when they result in different interpretations.

CHAPTER TWO

QUALITY OF DECISION

In comparing the nature and quality of the group decisions reached in the face to face vs. computerized conferencing modes of communication, we will look first at the ranking problem, "Arctic". This problem has a correct or criterion solution, and generated a great deal of quantified data related to the impact of the discussions on the decisions made. We will then turn to the more qualitative human relations problem, "Forest Ranger".

Quality Measures for the Arctic Problem

In the ranking problem, the procedures established by the originators of this task were followed. First, each individual read the problem in a room by himself or herself, and recorded in writing an initial opinion or decision on the problem. This decision was a rank ordering of the relative importance for survival in the arctic of fifteen items. In the face to face condition, the subjects brought their written rank orders with them to the conference room. In the computerized conference, they were given a special command ("share ranks") which would produce a table of their rank order in the conference for the others to see. After discussion, the subjects each gave their perception of the group's decision or ranking of the items, and their own final opinion as a result of the discussion.

The problem has a "correct" solution, or criterion, set by a group of "experts": the men and officers of the "Para Rescue Specialists, 413

Transport and Rescue Squadron, Canadian Forces Base, Prince Edward Island, Canada" (see Eady and Lafferty). These are the people who are trained and experienced in sub-arctic survival. Their group decision was used as the criterion. Interestingly, another group of "experts", four eskimos who live in the area described in the problem, also gave their answers, and they were very similar to the rank order established by the military survival experts.

Given these data, we can compute several kinds of deviation scores from the criterion or among scores reported by an individual. For example, one can compute the deviation score between the criterion and the individual's pre-discussion ranking. In this calculation, the raw (not squared) deviations are used, and whether an item was higher or lower is not taken into account, just absolute difference. If the expert group's ranking of rope was 2 and an individual ranked it 5, the difference would be 3. The "deviation score" for an individual is simply the sum of the deviations for each of the fifteen items ranked.

We can thus calculate the following kinds of means or averages

1. Individual deviation score- criterion= the quality of the individual's decision before discussion

2. For a group of five, the mean of the above five figures gives the average deviation score before discussion, or how good the group was before discussion.

Individuals and groups varied tremendously in the amount of knowledge which they brought to the situation. There are many facts which a person may or may not have known. For instance, one of the items on the list is a compass. The facts are that close to magnetic north and in proximity to iron ore deposits in the area, a compass is completely unreliable and useless. Some groups did not include any individuals who knew these facts, and some groups included several who knew this. Thus analyses must look at improvements or degradations in quality of solution (deviation from criterion), rather than simply at the absolute quality of the group's post-discussion ranking.

3. The post-discussion "Group" score was computed by taking the sum of the deviations between the criterion and the reported group decision for each individual. It should be noted that even in groups that thought they reached perfect consensus, this "group decision" may have been slightly different for each individual; in any case, it is their perception of the group decision. The mean of the sum of these five deviation scores is the "group score".

4. We can then look at the difference in quality between the pre-discussion individual scores and the post-discussion group scores. This can be done in raw or absolute terms, using various measures of a percentage improvement. It can also be done by analysis of covariance.

5. We can also look at amount of conformity or "commitment", measured as the difference between the individual final ranking reported by

each subject, and the group ranking reported by each individual.

6. We can look at individual opinion change, measured as the difference between an individual's pre-discussion rankings and his or her post-discussion rankings.

Summary of Findings

We tried several different ways of measuring improvement in decision quality and several different methods of analysis of variance. All results show the same thing:

1. There is no significant difference in amount of improvement in quality of decision between face to face and computer mediated discussions.

2. The quality of group decision improved about 25% after discussion, as compared to the average of the individual decisions before discussion.

3. There is no significant difference between media in amount of opinion change due to discussion.

4. There is no significant difference between media in conformity to group decision after discussion.

5. In most cases, the group decision was better than that of the best member before discussion. This was true of both communications modes.

Our first analysis was an analysis of covariance, shown in figure 2-1. Here, we are holding constant or covarying out the quality of the individual decision before ranking. The independent variables are mode of communication and order (whether this was the first problem the group had to solve, or the second.) There are no significant differences associated with either independent variable.

FIGURE 2-1
Analysis of Covariance:
No difference in Quality of Solution Reached
by Face to Face and Computerized Conferencing Media
Covariate= Quality of the Individual's Pre-Discussion Solution

A. Quality of Final Group Decision
Group Rank - Criterion
A = FtF vs CC, B = Problem Order

Source	df	F
A	1	.503342
B	1	.155436
AXB	1	.003405 (Not significant)

B. Quality of Individuals' Final Decision
Individuals Finals - Criterion

Source	df	F
A	1	1.164417
B	1	.244955
AXB	1	.004045 (Not significant)

N = 8 groups per condition; 5 subjects per group

Percentage Improvement

The following analysis gives more detail of what is occurring to produce "no difference". It is not that there is no improvement, regardless of what mode of discussion occurs, but rather that either mode of communication results in substantial improvement in the quality of decision.

In the first analysis of percentage improvement in Figure 2-2, we are looking at changes in the scores reported by the five individual members of each group as their initial decision and their perception of the group decision. A 2x2x4 nested design for analysis of variance was performed on these data, and showed no significant difference associated with mode of communication or order.

Though not statistically significant, there is a tendency for groups which had their arctic problem second to improve a little more.

Several other ways of computing percentage improvement were used, such as Individual deviation minus Group deviation divided by the maximum possible deviation. They all showed the same sort of effect, and none of the differences were significant when analyzed by a nested design for analysis of variance.

A second method of analysis of percentage improvement uses the group as the unit. It averages the five individual rankings before discussion to get the group's initial average ranking. Then it

averages the group member's five reported rankings for the "group decision" after discussion to get the "group" scores. We see the exact same pattern. Performance improves about 25% in either mode of communication.

Figure 2-2; Percentage Improvement by Mode of Communication

A. INDIVIDUAL IMPROVEMENT (80 Scores)

A = Mode of Communication

B = Order - B1 = Arctic first, B2 = Arctic second task

(Individual deviation minus Group deviation divided by Individual deviation) X 100.

	MEANS		
	FTF	CC	
first	23.3430	23.1350	23.2390
second	31.3495	25.5730	28.4612
	27.3462	24.3540	

B. AVERAGE RANKS OF FIVE GROUP MEMBERS (16 Scores)

Mean Percentage Improvement in Deviation from Criterion

	ftf	cc	
first	17.08	23.20	20.14
second	28.68	25.43	27.055
	22.88	24.315	

Individual vs. Group Scores

To what extent is the average 25% improvement of the group decision over the individual decisions attributable to

a) The group approaching agreement with the best member, vs

b) The group exhibiting a "collective intelligence", by pooling its knowledge and producing a better decision than any one of its individual members.

The data are shown in table 2-3. The Sandler's A statistic was not significant at the .05 level, and the comparisons between the CC and the FtF conditions also showed no significant difference. We are limited in our confidence in generalizing our observations because with only sixteen observations there must be very, very strong differences before they reach statistical significance. However, the data does suggest that "collective intelligence" process is the better explanation of the observed improvement in quality of decision due to discussion. Eleven of the sixteen groups produced better decisions than any of their members, and a twelfth was equal to its best member. When broken down by mode, two of the groups which produced poorer quality than their best member were in CC, and two in FtF. It is also interesting that both the most spectacular gain (+25.6) and the worst decline (-21.2) were in the CC condition.

Even approaching the decision of its best member is a desirable outcome for a group process. This is because there is usually no way of knowing before a discussion takes place which member indeed has the best solution.

Since "collective intelligence" is a phenomenon which interests us, we decided to look in more detail at the four cases where the group solution was not at least equal to that of the best member. We find that three of the groups, including the worst case, are characterized by the best member getting worse as a result of the discussion. This probably means they were not terribly committed to their initial views. The solution for the fourth group was the best of the four and better than the average for the 16 groups as a whole. In this case, the improvement of the best member and the deviation of the groups solution from the best member's initial solution was within 6%. In the remaining 12 groups, where the group solution was better or equal to the best member's solution, we find that in eleven groups the best member improved, and in one the best member stayed the same. In this latter case, another one of the members showed improvement. Of the eleven groups that did better than their initial best member, a "better best member" emerged in nine. This confirms the assumption that there was an initial distribution of knowledge among at least several different members of most groups, which contributed to obtaining an improved solution.

Figure 2-3
BEST INDIVIDUAL PRE-DISCUSSION SCORE VS. GROUP SCORE

Exp	Best Ind	Group	Diff
1a	42	44.0	-2.0
1b	24	18.0	6.0
1c	38	38.0	0
1d	58	50.0	8.0
2a	46	44.4	1.6
2b	66	64.0	2.0
2c	56	30.4	25.6
2d	56	51.6	4.4
3a	40	39.2	.8
3b	26	19.6	6.4
3c	48	41.6	6.4
3d	52	50.0	2.0
4a	28	30.0	-2.0
4b	52	73.2	-21.2
4c	50	64.0	-14.0
4d	50	42.4	7.6

Sum Diffs= 31.6
Sum of the Differences squared= 1579.04
Sandler's A= 1.579.04
Not significant

Figures 2-4 and 2-5 indicate no difference in opinion change or in conformity of individual to group decisions between media. There is a strong tendency for the group decision to be closer to the individual's final ranking in face to face discussions, but this is significant only at the .10 level.

Figure 2-4
Conformity by Mode of Communication
2X2X4 nested factorial

Group Rank - Final Individual Rank

Means

		Mode of Communication		
		FTF	CC	
Order of Problem	1st	11.90	18.5	15.20
	2nd	10.90	12.60	11.70
		11.40	15.55	

Nested Design

Source	SS	df	MS	F
A	344.45	1	344.45	3.3264
B	238.05	1	238.05	2.2989
A X B	120.05	1	120.05	1.1593
C/AB	1242.60	12	103.55	.8329
S/ABC	7956.80	64	124.325	
Total	9901.95	79		

Table Value for F
1 and 12 df = 4.75
12 and 64 df = 1.90

Pooled ANOV

Source	SS	df	MS	F
A	344.45	1	344.45	2.8456
B	238.05	1	238.05	1.9666
A X B	120.05	1	120.05	.9918
WG	9199.40	76	121.0447	
Total	9901.95	79		

Table Value for F
1 and 76 df = 3.97

Not Significant

A = mode
B = order
C/AB error term for A, B, A x B
S/ABC = error term for C/AB
WG = pooled error term

Figure 2-5: Opinion Change by Mode and Order
2X2X4 nested factorial

Individual rank - Final rank

A = Mode of communication
B = Order - B1 = Arctic first, B2 = Arctic second task

MEANS

	FTF	CC	
B1	39.50	41.30	40.40
B2	39.30	35.50	37.40
	39.40	38.40	

Source	SS	df	MS	F
A	20.00	1	20.00	.0876
B	180.00	1	180.00	.7888
A X B	156.80	1	156.80	.6871
C/AB	2738.40	12	228.20	1.0402
S/ABC	14040.00	64	219.3750	
Total	17135.20	79		

F = 4.75 for p = .05 and 1 and 12 df.
F = 1.90 for p = .05 for 12 and 64 df.

Since F for C/AB not significant we can pool error terms (that is combine
C/AB with S/ABC)

Source	SS	df	MS	F
A	20.00	1	20.00	.0906
B	180.00	1	180.00	.8153
A X B	156.80	1	156.80	.7102

WG	16778.40	76	220.7684
Total	17135.20	79	

F = 3.97 for p = .05 for 1 and 76 df.

No significant differences are obtained for either analysis.

Quality of Decision for Forest Ranger

We attempted to establish an overall criterion for the quality of decision reached by the groups in the "Forest Ranger" human relations problem. This involved the use of "experts" to rate the decisions. The experts were faculty members in the NJIT Department of Organizational and Social Sciences who have expertise in personnel matters.

First, the decisions actually reached by consensus or by disagreeing individuals were summarized in a paragraph or a few sentences, by examining the transcripts or listening to the tapes. These decisions were paraphrased in the actual words of the participants. We asked the judges to rank-order the quality of the decisions, on the following criteria:

1. Did the group accomplish its assigned task of actually making a decision?

2. Is the decision feasible, given the description in the problem of any resources or limitations that are available?

3. Likelihood of the decision leading to an effective outcome, both in the short term and in the long term.

4. Completeness of the decision.

A second approach was to extract the "decision atoms" from the complete decision, in terms of the actions recommended. Having identified the decision atoms, we could tabulate their frequency by communication condition, and also ask the judges to rank the atoms in isolation. These "atoms" and the accompanying instructions are shown in figure 2-6.

The first expert rating approach, having the judges rate the relative quality of the entire decision reached by a group consensus, or a majority or minority faction, failed to produce any consensus at all among the judges. For example, in rank ordering the solutions, the rank could vary between 1 and 24. The following complete decision was ranked as 19, 2, and 2 by the three judges:

"Evans should have a meeting with Bill and Joe and ask them what the problem is and why they are behaving the way they are. He should ask them how they feel they can change and still remain true to what they want. They should discuss their problems honestly... Evans main goal should be to try and get Bill and Joe to work together, using the advantages of each."

The following solution received ranks of 3,9, and 13:

"Joe should be made foreman and Bill should be a member of the crew. Bill should still get foreman's pay, and be trained by Joe."

The difficulty is that a total decision has many elements in it, and the judges differed about as much as the subjects did about some of

the elements.

Our second approach, issued simultaneously to the judges, was to isolate the distinct decision "atoms" or elements that composed the complete decisions. These are shown in the following table. The notation "B1" means a decision with respect to Bill, for example. The ratings of the individual decision atoms were requested on a one to ten scale.

Figure 2-6

ATOMS OF SOLUTIONS TO FOREST RANGER PROBLEM

The actual decisions to the "Forest Ranger" problem reached by various groups in the experiment are listed below. Considering each element individually, please rate them from "1" (first choice, best action that could be taken in this situation) to "10" (very poor decision; will have adverse consequences).

ACTIONS WITH RESPECT TO BILL

B1	Reinforce Authority
B2	Express Confidence
B3	Maintain as foreman
B4	Request to make compromise
B5	Request to take training
B6	Order to take training
B7	Weaken Authority
B8	Maintain pay
B9	Reduce pay
B10	Request to step down
B11	Make Co-foreman
B12	Make vice foreman
B13	Give another Job
B14	Demote for one year
B15	Demote indefinitely
B16	Make member of crew

ACTIONS WITH RESPECT TO JOE

J1	Make supervisor of foreman
J2	Make foreman
J3	Make co-foreman
J4	Make vice foreman
J5	Give raise
J6	Give Appreciation
J7	No salary change
J8	Request to train Bill
J9	Maintain current job
J10	Request to compromise, change attitude, reprimand
J11	Order to train Bill
J12	Threaten to fire

ACTIONS WITH RESPECT TO EVANS

E1	Take over authority, foremans job
E2	Appoint third party foreman
E3	Bring in outside expert
E4	Act as mediator, meet with principles
E5	Meet with crew
E6	Get view of crew
E7	Let crew decide
E8	Admit mistake

Though we can establish no correct or criterion solution for total decisions on the Forest Ranger problem, we can look at the decision components in terms of qualitative differences in the nature of the decision made and see if they differ between the media. There does seem to be a tendency for the computerized conferencing mode to be a little more positive or generous and less punitive in the decision reached. This conclusion is suggested by a content analysis of the specific decisions reached and their frequencies.

The table which follows shows the popularity of the various decision atoms for the unanimous face to face groups, the majority in CC groups, the minority in CC groups, and the expert judges. There were three judges, and the number from zero to three in the last column shows how many of them placed the decision component in their top five.

The table includes those decision atoms which were included as final decisions either in at least three FtF groups, or at least three CC groups (majority or minority components), or for which all three judges gave a top five rating.

One reads the table as follows. The decision atom "give Joe a raise" was included in all eight CC final group decisions. Since seven of the eight CC groups had both majority and minority (dissenting) opinions, one can compare majority to minority views in these groups.

Giving Joe a raise was included as one of the components in seven of the eight CC majority decisons, and in four of the seven CC minority decisions. Five of the eight FtF groups, all of which reached consensus, included this component. Finally, two of the three experts rated this component in their top five.

So far, then, we see that the reward-oriented option of giving Joe a raise is slightly more popular in the CC groups. It was the most frequent decision atom for CC groups, but not for FtF groups. There, the most popular component was punishment-oriented, that of reducing Bill's pay. Meanwhile, in five of the seven CC groups with minority holdouts, the minority refused to go along with this, and opted for maintaining Bill's pay. Looking to the experts, this is one of the few things that they are unanimous about-- that maintaining Bill's pay is correct, and that reducing his pay while maintaining him as an employee is a poor decision.

The only other options that received unanimous approval by the judges were not very popular among the subjects. Two of these three were democratically or reward oriented options: getting the view of the crew, and expressing confidence in Bill. These options appeared only in CC minority decisions.

There are two kinds of conclusions or speculations that we would like to make. One is that a total consensus is not always a "good" thing. As we will examine in the next chapter in detail, face to face groups are much more likely to be able to generate a total consensus, or put another way, to force minority-view members to stop pressing their

point of view and go along with the group. However, we have seen here that the minority points of view in the CC groups often tended to be "better" decisions, as rated by the expert judges.

Our second observation is purely speculative. Face to face groups have been observed to make more risky or extreme decisions than the individuals comprising them would make on their own. This has something to do with social-psychological pressures generated in face to face groups, and/or with the personality characteristics of the persons who tend to dominate face to face discussions. Perhaps the CC environment does not generate these pressures. We will see in the chapter on equality of participation and dominance that all of the face to face groups that decided to reduce Bill's pay had a member who contributed 30% or more of the interaction units, whereas the two FtF groups that decided in favor of maintaining Bill's pay did not have a dominant member.

Figure 2-7

DECISION OPTIONS VERSUS MODE

FOR

FOREST RANGER

DECISION ATOM BY NUMBER OF GROUPS OR EXPERTS	CC TOTAL	CC MAJORITY	CC MINORITY	FTF CONSENSUS	EXPERTS
GIVE JOE RAISE	8	7	4	5	2
MAINTAIN BILL'S PAY	7	3	5	2	3
REDUCE BILL'S PAY	7	5	2	6	0
MAKE JOE FOREMAN	6	5	2	4	1
MAINTAIN JOE'S JOB	5	1	4	3	2
MAINTAIN BILL AS FOREMAN	4	3	2	4	2
MAINTAIN JOE'S PAY	4	1	3	2	0
GIVE BILL ANOTHER JOB	3	2	2	0	1
REQUEST JOE TO COMPROMISE	2	2	0	4	1
WEAKEN BILL'S AUTHORITY	2	1	1	3	0
DEMOTE BILL FOR ONE YEAR	2	2	0	3	0
GET VIEW OF CREW	1	0	1	0	3
EXPRESS CONFIDENCE IN BILL	1	0	1	0	3
MANAGER ACTS AS MEDIATOR	0	0	0	1	3

CHAPTER THREE

ABILITY TO REACH CONSENSUS

For the ranking problem, consensus was measured by using Kendall's coefficient of concordance for the five "final group rankings" reported by each individual in each group. This varies from 0 for no agreement to 1.00 for perfect agreement on the placing of the fifteen items ranked by the group. The results are shown in Figure 3-1. There is a statistically significant difference in favor of face to face groups. However, substantively, the difference is not very large. All CC groups reached a reasonable amount of agreement. Some of those groups that did not reach near-total agreement seem to have run out of time; whereas all face to face groups completed their task within the 90 minutes allowed, many of the CC groups were cut off before they were able to finish. However, this is not the only factor. The computerized conference seems to provide little opportunity for a dominant leader to emerge to force a consensus, and an environment that is psychologically and socially more conducive to allowing persons to refuse to go along with the group when they think their decisions are better than those of the rest of the group members.

An interesting sidelight is that all of the face to face groups apparently THOUGHT that they had reached total consensus. However, in half of the groups, when individual members were asked to report this agreed upon decision in writing after the meeting, their versions of the decision were somewhat different. This is despite the

fact that the participants usually wrote down the supposed decision on a list of the items they had with them in the conference room, and later referred to it in reporting the decision.

Figure 3-1

Group Consensus on the Ranking Problem, by Medium of Communication

Face to Face	CC
.9897	.9774
1 .00	.8626
.9886	.9031
1 .00	.9857
.9943	.9671
.9989	.9811
1 .00	.9737
1 .00	.8077

Mann-Whitney U test

Ub=0

p< .01

Note: 1 .00 means perfect consensus, all five participants on all 15 items ranked

The difference in ability to reach consensus was much greater for the seemingly simpler, but amorphous and value laden, human relations problem. (See Figure 3-2). Consensus here was coded by simply looking at or listening to the final opinion given by each member, and seeing if there was agreement. Only one of the eight CC groups reached consensus on this problem, according to the transcripts, whereas all of the face to face groups reported reaching consensus.

We think that part of the difference is that an announced consensus in the face to face groups may have in fact not been present. Unlike the procedure followed for the complex ranking problem, the members of the group were not required to explicitly state what the "group decision" was or whether they agreed with it. It is quite likely that in at least some of the groups, there were persons who did not agree with the decision announced by a person playing a leadership role, but who chose not to make their disagreement explicit.

However, most of the apparent differences are probably related to aspects of the nature of the two tasks and the structuring of the interaction processes used in the tasks.

Figure 3-2

Inability of Computerized Conferencing Groups to Reach Consensus on Unstructured Problems

Mode	Consensus	No Consensus	Total
Computerized	1	7	8
Face to Face	8	0	8
Total	9	7	16

Chi square=3.06, $p<.05$

We have seen that the results of comparing computerized conferencing and face to face discussions on ability to reach consensus depended somewhat upon which of the two problems was being discussed. We initially thought of the problems as different in the sense that the Forest ranger was "a simple human relations problem" and that Lost in the Arctic is a complex, scientific type task with a correct answer. However, there are other differences evident between the two. We use a correlated T-test to compare the questionnaire answers for the two problems, ignoring mode of discussion.

Lost in the Arctic, though a complex and somewhat difficult task, is more interesting, and much more structured. It is clearer to the participants what they must do, and easier for them to systematically attack and complete the problem.

The results of the T test for differences between the problems show that the mean rating for degree of interest was better for Arctic. (Mean for Forest Ranger was 2.8 and for Lost in the Arctic, 2.2 on a one to seven scale where 1 is completely interesting.) ($T=3.73$, $p=0.00$). The issues involved were also much clearer (Mean for Artic, 2.2; for Forest Ranger, 2.8; $T=3.18$, $p= 0.00$).

Typically, a group attacked Arctic by comparing their 15 initial rankings and then picking out a subset of items near the top; then agreeing first what would be number one, then number two, etc. With

Forest Ranger, it seemed to be much more difficult for a group to know where to begin, and how to focus their discussion. Our feeling is that an unstructured problem needs strong human leadership to structure the discussion and decision making process; computerized conferencing doesn't appear to facilitate the natural emergence of a leader.

The Importance of Sociability

There are indications that social-emotional content is crucial for a group's effective functioning in this medium. It seems to provide the necessary motivation and cohesion for cooperation in task orientation.

The transcript of the training session for the eight groups which subsequently solved the ranking problem via computerized conferencing were Bales coded by a single assistant (Thus, the reliability is unknown; we did not invest the resources to double code all transcripts because this is an exploratory analysis, on a relationship hypothesized ex post facto, rather than before the experiments were conducted). In Figure 3-3 are the results for the numbers of positive comments (Bales categories 1,2, and 3, showing social solidarity, showing tension release, and agreeing) during the training session. In the second column is the Kendall's coefficient for the degree of agreement reached by that group one week later, when it was given its tasks. The groups were rank ordered on the relative number of social-emotional positive comments sent during the training, and the amount of consensus reached. Rho was used as a measure of association, and tables of significant values for Rho consulted to see if the rho was significant with an N of eight groups. The rho of .898 is significant at the .01 level.

Rho was also computed for the relationship between the number of social-emotional comments-NEGATIVE during training, and subsequent degree of consensus. The rho of .285 was not significant.

Ideally, one would test the relative importance of "social-emotional positive" comments in face to face vs. CC by doing a similar analysis for the groups which solved the ranking problem face to face. However, all of these groups reached complete (100%), or near complete (98-99%) agreement. Thus, our dependent variable (degree of consensus) is not able to distinguish among them. Put another way, it does not seem to matter how much solidarity they established the first week, face to face groups were always able to achieve consensus in week two. Why this occurs will be explored further in the chapters on interaction process.

Figure 3-3: Number of Social-Emotional Positive Units vs. Degree of Consensus

Group	#Bales positive	rank	Kendalls	rank
IA	58	6	.9774	6
ID	37	1.5	.8626	2
2C	51	4	.9031	3
2d	65	7	.9851	8
3C	52	5	.9671	4
3D	114	8	.9811	7
4b	38	3	.9737	5
4D	37	1.5	.8077	1

Rho=.898, $p<.01$

Sex, Medium of Communication, and Opinion Change

Another piece of the puzzle that explains why CC groups were more likely to fail to reach consensus is that the females in a group are less likely to change their opinions in the direction of the opinions held by males. Opinion change was operationally defined as deviation scores between an individual's pre-discussion ranking and the post discussion ranking they reported as their own opinion at that time (as compared to their reported impression of the group's ranking).

A Mann-Whitney test was used on the Z scores. It was hypothesized that in the face to face condition, females would change their opinions more than males. This was significant at the .01 level. In the CC condition, there was no significant difference between males and females in the amount of opinion change.

Dominance and Consensus

Finally, we suspect that one of the most important factors is that CC as a mode of communication is not conducive to the spontaneous emergence of a dominant group member, or leader. Especially in an unstructured, value-laden task such as the Forest Ranger problem, we think that leadership is very highly correlated to the probability of obtaining a group consensus. This hypothesis will be examined in the chapter on equality of participation and dominance.

Finally, we wish to reiterate that consensus is not a particularly necessary or always a good goal for a group to achieve. In the case of the Arctic problem, the average of the decisions in the non-consensus groups was just as good as the group decision in consensus groups. This is shown in Figure 3-4 as a very low correlation between our measure of consensus and our measure of quality of decision. And in the Forest Ranger problem, it will be remembered from the preceeding chapter, the decision components or atoms that distinguished the minority opinions in CC tended to be highly rated by the expert judges which we used.

Figure 3-4
Correlation between Kendall's and Group rank - Criterion.

Group	Kendall's	Rank	Ave. Deviat	rank
1a	.9774	6	44	8
1b	.9897	11	18	16
1c	1.000	14.5	38	12
1d	.8626	2	50	5.5
2a	.9886	9	44.4	7
2b	1.000	14.5	64	2.5
2c	.9031	3	30.4	13
2d	.9857	8	51.6	4
3a	.9943	12	39.2	11
3b	.9989	10	19.6	15
3c	.9671	4	41.6	10
3d	.9811	7	50	5.5
4a	1.000	14.5	30	14
4b	.9737	5	73.2	1
4c	1.000	14.5	64	2.5
4d	.8077	1	42.4	9

Spearman's Rho = .1098

Ranked deviation score 1 = largest deviation (poorest decision).
Ranked Kendall's 1 = Lowest Kendall's (least agreement).

Chapter 4

DIFFERENCES IN COMMUNICATIONS PROCESS, I: AMOUNT AND TYPE OF COMMUNICATION

The main method for quantifying the communications process used in this experiment is Bales' Interaction Process Analysis (IPA). This technique breaks all communications into units, which are the equivalent of a simple sentence or a single thought. Each unit is then coded into one of twelve mutually exclusive categories, i.e., agrees, disagrees, gives opinion, asks for opinion, etc. We will first review the procedures used in creating the IPA data for this study. Then we will look at the results for

1) Differences in IPA distributions between the FtF and CC modes of communication, when problem (task type) is held constant

2) Differences in IPA distributions between the problems, when communication mode is held constant

3) Differences in amount of communication between the FtF and CC modes, as measured by Bales units.

4) Relationship between communications process as measured by IPA and communications outcome in terms of consensus and quality of solution. This analysis can be done only for the Arctic problem.

Procedures for Bales IPA Coding

The IPA coding for this experiment compares communications in the audio channel for the FtF conferences with the written channel in CC. It excludes all totally non-verbal communication (facial expressions or gestures) in the FtF condition, and all of the non-communicated verbal and non-verbal expressions emitted by individuals at their terminals in the CC condition.

The coding for the Interaction Process Analysis was done in the following manner:

1. The coders read Bales' book on Interaction Process Analysis, including the appendix.

2. Coders were trained as a group; then practiced in pairs until they achieved reasonable consistency. Their first coding was checked unit by unit and they started coding in an unsupervised manner only after their coding was found to match the coding standards established for the group to follow.

3. CC transcripts were independently coded by two coders. They then met to review the entire transcript and resolve any inconsistencies. If they were unable to decide on a coding difference, they consulted the study director.

4. The tapes were listened to simultaneously by two coders. They

agreed on the start and end of the units heard and on the coding for each unit. The FtF conferences were recorded with a separate microphone and tape track for each participant, so that the speaker could easily be identified when the tapes were played back. With this method, the speaker being coded sounded loud and distinct and was easily identifiable, while the inputs from the other speakers were soft but audible and provided the coding context.

While production of a written transcript from the tapes and their independent coding by two coders might have been preferable, this was too time consuming and expensive. As it was, coding the data took many times longer than actually running the experiments.

The number of Bales units per face to face group was much greater than the number for a CC group. Therefore, each individual and group was transformed to a percentage distribution among the twelve categories. Then statistical tests were performed to determine if there were any significant differences in IPA distributions associated with mode of communication, problem, order of problem, and the interaction among these variables in relation to the percentage distribution for each of the Bales categories.

There are many different ways in which the percentages could be computed. To take full advantage of the design, we computed the percentage distribution for each individual, in each condition. Thus, we actually have the Bales distributions for each of 80 individuals in a face to face conference, and in a computerized conference.

The mode of analysis was the two by two factorial nested design explained above. If there was no significant group effect, then the error terms could be "pooled", meaning we could use the 80 observations as independent observations for statistical test purposes. We also performed a non-parametric test on the data for each Bales category, which gave us similar results, but did not turn up as many statistically significant differences in IPA distributions, since it is a less powerful analytic tool.

Differences Associated With Communication Mode

The detailed analysis of variance tables are included as an Appendix. Note that the analyses were first performed separately for the two problems, using communication mode as the independent variable. For each problem, we tested the significance of mode of communication, order (whether it was the first or second problem solved by the group), and the interaction between mode and order.

Listed in figures 4-1 and 4-2 is a summary of the statistical results of the 24 analyses of variance. The first two columns show the mean percentage of communications in each category. For example, in the first table, results for Forest Ranger, the first column shows that on the average less than 1% of an individual's communications were verbally "showing solidarity", but in CC, 3.22% fell into this category. The third column shows that the results for the 16 groups in the nested factorial design were significant at the .005 level, meaning that the probability of the observed differences occurring by chance in a sample this size is one in 200. The fourth column shows

the level of significance if the group was not a significant variable and the observations could be pooled, with the 80 individuals treated as independent observations. In this case, group was significant, so the pooled analysis could not be done. Finally, the last two columns note if there were any significant differences associated with the order of the modes (whether the face to face discussion was first or second), or with the interaction between mode and order.

In looking at these data , there is an apparent coding problem. Even for the Forest Ranger problem, face to face, we obtained a somewhat different distribution of coding than did persons coding problem discussions such as this who were directly trained by Bales. (See Bales and Borgatta, 1955, p. 400 for the complete distributions). Our coding has 20% more of the statements classified as "giving opinions" than Bales and Borgatta code, and correspondingly lower percentages in all of the other categories. This means that our results cannot be directly compared to those of other investigators, since apparently the training for coding interpreted many more statements as representing some sort of analysis or opinion than "should" be there, according to the distributions obtained for similar studies by Bales and his colleagues. Other possible explanations for the coding distributions obtained are

1) The non-verbal content coded in other Bales studies tends to be heavily concentrated in the social-emotional categories. Since we did not code this, our resulting distributions will of course be different.

2) The "practice" effect of communicating as a group on CC before receiving a problem to solve may have affected their communications even in subsequent FtF conferences.

3) Perhaps Upsala College has produced an unusually opinionated and analytic set of students, compared to the subjects used in other studies.

The skewed coding distributions do not affect the comparisons among problems and modes for this study, since all of the coders were coding the data with the same guidelines and interpretations. In the majority of cases, the same pair of coders coded both the CC and FtF condition for the same group. In any case, the seven individuals who did the coding had been trained to an acceptable level of inter-coder reliability.

Figure 4-1

SUMMARY OF IPA RESULTS FOR

FOREST RANGER BY

MODE OF COMMUNICATION AND ORDER

BALES CATEGORY	AVERAGE		P SIGNIFICANCE		P SIGNIFICANCE	
	FtF	CC	BY GROUP	POOLED	BY GROUP	POOLED
SHOWS:						
SOLIDARITY	.79	3.22	.005	GS		
TENSION RELEASE	3.98	.83	.0005	.0005		
AGREEMENT	13.19	4.79	.0005	.0005		
GIVES:						
SUGGESTIONS	4.70	9.21	.10	.10		
OPINION	54.21	53.92	X	X		
ORIENTATION	12.81	16.10	.10	.02		
ASKS FOR:						
ORIENTATION	3.27	1.58	.05	GS		
SUGGESTIONS	.30	.62	.25	.20		
SHOWS:						
DISAGREEMENT	4.85	2.39	.05	.05		
TENSION:	.81	2.16	.05	.01	ORDER:	
PROBLEM 1ST	.28	1.68			X	.05
PROBLEM 2ND	1.33	2.64				
ANTAGONISM:	.75	1.67	X	X		

GS = GROUP SIGNIFICANT CANNOT POOL BY INDIVIDUAL

Figure 4-2

SUMMARY OF IPA RESULTS FOR

ARCTIC BY

MODE OF COMMUNICATION AND ORDER

BALES CATEGORY	AVERAGE		P SIGNIFICANCE		P SIGNIFICANCE	
	FtF	CC	BY GROUP	POOLED	BY GROUP	POOLED
SHOWS:						
SOLIDARITY	1.66	2.44	.10	.05		
TENSION RELEASE	7.70	1.60	.0005	.0005		
AGREEMENT	13.35	6.82	.01	GS		
GIVES:						
SUGGESTIONS	3.56	4.89	.20	.10	MODE X ORDER	
PROBLEM 1ST	2.95	6.17			.025	.02
PROBLEM 2ND	4.17	3.61				
OPINION	42.99	57.80	.005	GS		
ORIENTATION	14.58	11.81	.25	GS		
ASKS FOR:						
ORIENTATION	3.72	1.62	.025	.0005		
SUGGESTIONS	1.14	.58	X	GS		
SHOWS:						
DISAGREEMENT	3.51	2.46	X	GS		
TENSION:	1.52	.64	.025	.005		
ANTAGONISM:	1.11	1.86	X	GS	ORDER:	
PROBLEM 1ST	.77	.73			.05	GS
PROBLEM 2ND	1.45	3.00				

GS = GROUP SIGNIFICANT, CANNOT POOL BY INDIVIDUAL

Figure 4-3

FOREST RANGER SUMMARY

CC GREATER THAN FtF:

Significant Difference Observed

Shows Solidarity (.005)
Asks For Opinion (.01)
Shows Tension (.01)
Gives Orientation (.02)

Potential Difference Observed

Gives Suggestion (.10)
Asks For Suggestion (.20)

CC AND FtF THE SAME:

Shows Antagonism
Gives Opinions

FtF GREATER THAN CC

Potential Difference Observed

None

Significant Difference Observed

Shows Disagreement (.05)
Asks For Orientation (.05)
Shows Agreement (.0005)
Shows Tension Release (.0005)

ORDER: More Showing of Tension in both modes when problem is second (.05).

Figure 4-4

ARCTIC SUMMARY

CC GREATER THAN FtF:

Significant Difference Observed

Gives Opinion (.005)
Shows Solidarity (.05)

Potential Difference Observed

Gives Suggestions (.10)
Asks For Opinion (.20)

CC AND FtF THE SAME:

Asks For Suggestions
Shows Disagreement
Showing Antagonism

FtF GREATER THAN CC:

Potential Difference Observed

Gives Orientation (.25)

Significant Difference Observed

Shows Agreement (.01)
Shows Tension (.005)
Asks For Orientation (.0005)
Shows Tension Release (.0005)

MODE X ORDER: More Suggestions Given in Artic when 2nd problem in FtF and when 1st in CC (.02).

ORDER: Higher Antagonism if Arctic 2nd (.05).

Discussion of the Results

The twelve categories in Bales' Interaction Process Analysis can be combined into four main functional areas. Categories 1-3 and 10-12 are the "social-emotional" functions, oriented towards internal group process. The first three are called "social-emotional positive", while 10-12 are "negative". Categories 7-9 are "Task oriented", giving answers or contributions to solving the problem faced by the group, and categories 4-6 are varieties of "asking questions" in the task oriented area.

It will be noted, by way of further introduction, that there are some very strong differences in the profiles, even in the same medium, depending upon the type of task faced by the group, and that there is some interaction between task type and medium. For example, more tension was shown in the Arctic problem in the CC condition; more in the Forest Ranger problem in the FtF condition. These differences associated with problem will be detailed subsequently.

We will take each of the categories, describing more fully what is included in them, and then discuss the extent to which there appear to be significant differences between the media in the relative prevalence of communications of that type. We will also try to explain the possible reasons for or implications of significant differences that are discovered.

1. "Shows solidarity, raises other's status, gives help, reward"

Included in this category are initial and responsive acts of active solidarity and affection, such as saying "hello" and making friendly or congenial remarks to "break the ice"; praising or encouraging the other(s); giving support or sympathy or offers of assistance; urging harmony and cooperation. These are all overt attempts to improve the solidarity of the group.

Note that there is a significantly greater amount of "showing solidarity" in computerized conferencing. This is probably because much of the behavior of this type in a face to face situation is non-verbal, such as smiling in a friendly manner while nodding encouragement. Non verbal acts in this category are not codable from the tapes of the discussions. In the CC condition, however, the participants realize that they must put such things into words.

Another possible explanation is that the greater tendency towards overt, explicit showing of solidarity is an attempt to compensate for the perceived coldness and impersonality of the medium.

2. "Shows Tension Release, jokes, laughs, shows satisfaction"

This includes expressions of pleasure or happiness, making friendly jokes or kidding remarks, laughing.

There was significantly more tension release overtly expressed in the face to face groups. Much of this was waves of laughter,

particularly in the Arctic problem. The participants did not put this into words in the conference when typing. Observing them, however, there was much private laughter and verbal expressions showing "tension release", but these do not appear in the transcript. It is part of the private "letting down of face" that occurs but is not communicated through the computer.

3. "Agrees, shows passive acceptance, understands, concurs, complies"

This occurs as concurrence in a proposed course of action or carrying out of any activity which has been requested by others. There is significantly more agreement overtly expressed in face to face conferences than in computerized conferences. We suspect that this is related to the pressure to conform created by non-verbal behavior and the physical presence of the other group members. In any case, it is undoubtedly related to the greater difficulty of CC groups in reaching total consensus.

4. "Gives suggestion, direction, implying autonomy for other"

Includes giving suggestions about the task or suggesting concrete actions in the near term to attain a group goal. There is a tendency for more suggestions to be given by more people in computerized conferencing. This is part of the equalitarian tendency for more members to actively participate in the task behavior of a group in CC. In one of the problems, the difference was statistically

significant at the .05 level; whereas in the other it was sizable but did not reach statistical significance.

5."Gives opinion, evaluation, analysis, expresses feeling, wish"

Includes all reasoning or expressions of evaluation or interpretation.

This is the most frequent type of communication for both problems and both modes. For the Bales problem, there was no difference in its prevalence associated with mode of communication. For the Arctic problem, however, there was a large and statistically significant difference, with more opinion giving in the CC condition.

6. "Gives orientation, information, repeats, clarifies, confirms"

This includes statements that are meant to secure the attention of the other (such as "There are two points I'd like to make..."), restating or reporting the essential content of what the group has read or said; non-inferential, descriptive generalizations or summaries of the situation facing the group. There are no clear differences here. Whereas there is a statistically significant difference in the direction of giving more orientation in CC for Forest Ranger, for the other problem the difference is reversed.

7. "Asks for orientation, information, repetition and confirmation"

There is a significant tendency for this to occur more often in face

to face discussions. This is probably because of the frequency with which a group member does not hear or understand the pronunciation of a sentence or partial utterance. In CC, people are usually more careful to state their thoughts clearly, and the recipient can read it several times (rather than asking for repetition) if it is not understood the first time or is later forgotten. We have noticed many CC participants going back and looking at comments a second or third time; in a face to face discussion, they would probably ask something like: "What was it you said before about x?".

8. "Asks for opinion, evaluation, analysis, expression of feeling"

This occurs more frequently in computerized conferencing. For one of the problems, the difference reached statistical significance, whereas it did not for the other. This tendency to more frequently and explicitly ask for the opinions of all the other group members, as well as to more spontaneously offer ones own opinions and analyses in CC, does seem to qualitatively be characteristic of the medium.

9. "Asks for suggestion, direction, possible ways of action"

This includes all overt, explicit requests, such as "What shall we do now?" It is not very prevalent in either medium, and there are no significant differences.

Comparing our results to Vallee et. al.'s (1974) prediction that a precise count would show more "asking questions" in face-to-face discussions than in CC, we find that it depends on what kind of

"question". Questions of fact or information are more frequent in FtF, but questions about the opinions of others more frequent in CC.

10. "Disagrees, shows passive rejection, formality, witholds resources"

This includes all the milder forms of disagreement or refusal to comply or reciprocate. This is also an infrequent form of communication, but it occurs more in face to face discussions than in CC.

11. "Shows tension, asks for help, withdraws out of field"

Includes indications that the subject feels anxious or frustrated, with no particular other group member as the focus of these negative feelings. The results on this are rather puzzling. We end up with a statistically significant tendency for there to be more tensions when in CC for the Forest Ranger problem, but in FtF for the Arctic problem. Substantively, the proportion of these communications is very small in any case, and therefore the small differences are not important.

12."Shows antagonism, deflates other's status, defends or asserts self"

This includes autocratic attempts to control or direct others, rejection or refusal of a request, deriding or criticizing others.

This is infrequent in both media and there are no significant differences.

Effects of Order

For the most part, it did not matter whether the CC or the FtF discussion was held first. However, more suggestions were offered on the Arctic problem if it was discussed in CC as the first problem, but more in FtF discussion if the FtF was preceded by a CC condition. This is consistent with the tendency for CC to promote more giving of suggestions; apparently, the tendency carries over to a subsequent face to face conversation. This raises the interesting possibility that the group process and structure can be permanently changed by the experience of interacting through CC, a change that will carry over even to communications in other modes. Other pieces of evidence from other studies, including self reports of participants in long term field trials, indicate the same possibility.

Figures 4-3 and 4-4 give a more qualitative summary of the significant results shown in the preceeding two tables. For Forest Ranger, the differences between FfF and CC were statistically significant in eight of the twelve IPA categories. For arctic, the differences were significant in six of the twelve. However, these six do not in all cases correspond to the same eight that were significant on the other problem. Comparing the specific differences observed, one sees that they are a product of our second independent variable, task type, as well as of mode of communication. It appears, however, that greater verbalization of behavior that shows

solidarity occurs in CC regardless of task type, whereas more overt verbalization of agreement and of tension release occurs face to face.

Differences in Type of Communication by Problem

The second set of analyses of variance using the percentage of each individual's communications within each of the twelve Bales IPA codes compared the differences in the distributions obtained for the two problems, holding the communication mode constant. These results are summarized in figures 4-5 through 4-8.

We do find some confirmation that we are dealing with two distinct types of tasks and/or communication structures, based on some significant differences in the distributions obtained. In the computerized conferencing discussions, there was significantly more agreement for the Arctic problem, and significantly more tension shown for the Forest Ranger problem. This would be in line with our characterization of the Forest Ranger problem as a value-laden one, and of the Arctic problem as a knowledge-pooling problem.

We have no theoretical basis for explaining the other differences observed. Task type is not the main focus of our interest in this experiment, and we do not have a thorough knowledge of the literature. By reporting the results, perhaps others will see an overall pattern or theoretical analysis that does not occur to us. The main point which we wish to offer as a generalization on the basis of these data is that communication behavior is most definitely a function of task type as well as of mode of communication. We also feel that it is a function of the particular structure imposed upon the mode of communication, a theorem that will be discussed more in

the chapter on inequality and in the final section of the report which gives our design for the next experiment in this series.

Figure 4-5
FACE-TO-FACE BY
PROBLEM AND ORDER

BALES CATEGORY	AVERAGE FTF	AVERAGE CC	P SIGNIFICANCE BY GROUP	P SIGNIFICANCE POOLED	P SIGNIFICANCE BY GROUP	P SIGNIFICANCE POOLED
SHOWS:						
SOLIDARITY	.79	1.66	.10	GS		
TENSION RELEASE	3.78	7.70	.01	.0005		
AGREEMENT	13.19	13.35	X	X		
GIVES:						
SUGGESTIONS	4.70	3.56	X	X	ORDER:	
PROBLEM 1ST	6.74	2.95			.10	.10
PROBLEM 2ND	2.66	4.17				
OPINION	52.74	42.99	.025	.005		
ORIENTATION	12.81	14.58	.25	.20		
ASKS FOR:						
ORIENTATION	3.27	3.72	X	X	ORDER:	
PROBLEM 1ST	2.84	3.13			.20	.20
PROBLEM 2ND	3.69	4.31				
OPINION	2.88	5.15	.025	.001		
SUGGESTIONS	.30	1.14	.10	GS		
SHOWS:						
DISAGREEMENT	3.79	3.51	X	X	MODE X ORDER:	
PROBLEM 1ST	2.73	4.18			X	.05
PROBLEM 2ND	4.85	2.84				
TENSION:	.81	1.52	.20	GS		
ANTAGONISM:	.75	1.11	X	GS		

GS = GROUP SIGNIFICANT, CANNOT POOL BY INDIVIDUAL

Figure 4-6

FACE TO FACE SUMMARY

FOREST RANGER GREATER THAN ARCTIC

Significant Difference Observed

Gives Opinion (.005)

Potential Difference Observed

None

FOREST RANGER AND ARCTIC THE SAME

Shows Agreement
Gives Suggestions
Asks For Orientation
Shows Disagreement
Shows Antagonism

ARCTIC GREATER THAN FOREST RANGER

Potential Difference Observed

Gives Orientation (.20)
Shows Tension (.20)
Asks Suggestions (.10)
Shows Solidarity (.10)

Significant Difference Observed

Ask Opinion (.001)
Shows Tension Release (.0005)

MODE X ORDER: More Disagreement when Forest Ranger second but less when Arctic second (.05).

MODE X ORDER: Giving Suggestions greater when Forest Ranger first but greater for Arctic when second (.10).

ORDER: Less asking for Orientation when Problem is second (.20).

Figure 4-7
COMPUTERIZED CONFERENCING BY
PROBLEM AND ORDER

BALES CATEGORY	AVERAGE FTF	AVERAGE CC	P SIGNIFICANCE BY GROUP	P SIGNIFICANCE POOLED	P SIGNIFICANCE BY GROUP	P SIGNIFICANCE POOLED
SHOWS:						
SOLIDARITY	3.22	2.44	.25	GS		
TENSION RELEASE	.83	1.60	.20	.20		
I59Q 4.79	6.82		.20	.05		
GIVES:						
SUGGESTIONS	9.21	4.89	.10	.10		
OPINION	52.28	57.80	.20	.10		
ORIENTATION	16.10	11.82	.10	GS		
ASKS FOR:						
ORIENTATION	1.58	1.62	X	GS		
OPINION	5.35	7.46	.20	.10		
SUGGESTIONS	.62	.58	X	GS		
SHOWS:						
DISAGREEMENT	2.17	2.46	X	X	ORDER:	
PROBLEM 1ST	1.95	1.76			.25	.20
PROBLEM 2ND	2.39	3.17				
TENSION:	2.16	.64	.025	.005		
ANTAGONISM:	1.67	1.87	X	X	MODE X ORDER:	
PROBLEM 1ST	1.95	.74			.20	.20
PROBLEM 2ND	1.38	3.00				

GS = GROUP SIGNIFICANT, CANNOT POOL BY INDIVIDUAL

Figure 4-8

COMPUTERIZED CONFERENCING SUMMARY

FOREST RANGER GREATER THAN ARCTIC

Significant Difference Observed

Shows Tension (.005)

Potential Difference Observed

Gives Suggestions (.10)
Gives Orientation (.10)
Shows Solidarity (.25)

FOREST RANGER AND ARCTIC THE SAME

Ask Orientation
Ask Suggestions
Disagrees
Shows Antagonism

ARCTIC GREATER THAN FOREST RANGER

Potential Difference Observed

Shows Tension Release (.20)
Ask Opinion (.10)
Gives Opinion (.10)

Significant Difference Observed

Shows Agreement (.05)

ORDER: Disagreement higher when problem second (.20).

MODE X ORDER: More Antagonism when Forest Ranger first and when Arctic second (.20).

Amount of Communication, By Medium and Problem Type

The Bales units make it possible to get a comparable measure of the amount of communication taking place in the two media.

We see in figure 4-9 displaying these data that, as in the earlier pilot studies, there is unquestionably more communication taking place during the same amount of elapsed time in a five person group that discusses a problem in a face to face conference than in a computerized conference. It is in the range of two to three times as many communication units. There is no need to do a significance test on these data, since there is no overlap whatsoever (all FtF groups have more units than all CC groups). However, a Mann-Whitney U test was performed, and the differences are significant at the .01 level.

The difference in number of units between the two problems is probably largely accounted for by the fact that groups had 60 minutes to solve "Forest Ranger", but 90 minutes for "Arctic". We can only speculate about why the ratio for amount of communication was even greater for the shorter-time, qualitative values problem (Forest Ranger) than for the longer time-limit, scientific ranking problem (Arctic). It may be that with the short practice period given in this experience, the first 30 minutes or so in the computerized conference saw individuals not yet "up to speed", and they were just getting the hang of discussing things via computer when the hour was up. This would be supported by the observation that an average of about 75 units per person for the two thirty minute periods in Forest

Ranger is a lower rate than an average of 161 units per person for the three half hour segments in the 90 minute Arctic discussion. We do not see this increase in throughput of units in comparing the face to face groups.. they averaged about 460 units per half hour for the 60 minute discussion and 411 units per half hour for the 90 minute discussion.

Another possibility is that the value-laden Forest Ranger problem elicited more inactive "think time" in CC, where individuals just sat quietly and thought about the issues and choices. In a face to face conference, silences are against the norm.

To summarize, face to face conferences seem to generate two to three times the amount of communication in the same length of time as a computerized conference. The ratio is apparently influenced by the nature of the problem being discussed, group size (which we did not explore in this experiment), and the length of the meeting or discussion.

Figure 4-9

Amount of Communication by Mode, Problem, and Order of Problem

Number of Bales Communication Units

Arctic Problem

Mode of Communication

Order of problem		FtF	CC
	1st	2056	568
		1307	529
		1063	464
		954	347
	Means	1345	477
	2nd	1595	506
		1049	497
		946	479
		896	472
	Means	1121	489
	Both	1233	483

Ratio of FtF to CC=2.32

Forest Ranger Problem

Order of problem		FtF	CC
	1st	1085	332
		1027	284
		989	261
		518	256
	Means	905	283
	2nd	1301	394
		947	316
		795	301
		659	269
	MEANS	925	320
	Both	915	302

Ratio of FtF to CC= 3.03

Figure 4-10

% of Group Communications in Each Bales Category, by Group Consensus, Quality of Group Decision, and Quality of Best Final Individual Solution

Spearman's Rho Correlations and Level of Significance

Category	Consensus (Kendall's)	Group Decision Quality	Best Final Individual Quality
1. Shows solidarity	-.058	.624	.347
2. Shows tension release, jokes	.622	-.460	-.545
3. Agrees	.766	-.078	-.268
4. Gives suggestions	-.133	.303	.213
5. Gives opinions	-.824	.224	.340
6. Gives orientation	.440	-.471	-.318
7. Asks for orientation	.692	-.287	-.288
8. Asks opinion	-.554	.158	-.123
9. Asks for suggestion	.654	.051	.083
10. Disagrees	.258	-.037	.243
11. Shows tension	.591	.024	.082
12. Shows antagonism	.025	.207	-.055

Critical values for Spearman's Rho by Level of Significance

.10=.425
.05= .506
.01= .665

The Relationship of Bales Distributions to Group Consensus and Quality of Decision

The Arctic problem allowed us to obtain a measure of the amount of group consensus on the final decision (Kendall's coefficient of concordance, which varies from 0.0 to 1.00) and of the quality of the group decision (Deviation between the criterion and the mean group report of the group decision, or ranking, for each item). We have shown that there was less consensus in CC, but no difference in quality of solution. This is despite the fact that there were 2.3 times more communication units in the same amount of time in FtF discussions of the Arctic problem. We were led qualitatively to the supposition that there must be something more efficient about the communication process in CC, in terms of the process creating improvements in group decision quality without as many communication units. In figure 4-10 we show the data on the differences between the media in the distribution of Bales units, and for the relationship between the percentage of units in each of the Bales categories to group consensus and quality of group decision. We can thus gain some insight into what it is about the communication processes in CC vs. FtF that produces the observed differences in consensus formation and the observed lack of difference in decision quality.

Up until now, we have been working with the differences in individual communication behavior, measured as the percentage of communication units for each individual in each category. For this analysis, we

will change our independent variable to the percentage of total group communication units (all five individuals) in each of the Bales categories. Spearman's Rho is used in Figure 4-10, since the Kendall's coefficients do not meet the assumptions of Pearson's R, and we want to compare the relative strengths of correlations for consensus and quality.

We find many sizable and/or statistically significant relationships which, when combined with the information on the differences between media, help us to understand the consequences of the media for the dependent variables, quality of decision and amount of consensus on the final group decision.

Some of the correlations are not surprising at all, and in fact help to validate the Bales coding. For example, there is a .766 correlation between the amount of showing agreement and the final ability of the group to reach consensus. This also indicates one of the processes which explains the lower consensus in CC groups, since they have significantly less "showing agreement" type statements.

"Showing tension release", such as joking and laughing, is more prevalent in face to face groups, we saw above. This has a very significant relationship to ability of the group to reach consensus. However, it also has a strong, significant NEGATIVE relationship to the quality of the group's decision. It makes everybody feel good, but seems to detract from the quality of the group's product. Thus, we see in these two categories that two of the types of communication which are more likely to occur in face to face groups than in CC

groups do lead to group consensus, but do not lead to high quality decisions. They are generally considered good kinds of communications to have lots of, because they feel good and help the group reach consensus, but they are not objectively "good" things to have too much of.

The strongest relationship that we see is a negative one between giving opinions and ability to reach decisions. There was significantly more giving of opinions in CC, and giving opinions seems to prevent the group from reaching consensus. However, giving opinions is positively related to the quality of the group decision reached. We see a similar pattern, though not as strong, for the obverse of this, asking for opinions. A similar pattern of significant differences appears for "giving orientation". It occurred more frequently in FtF groups. It has a significant positive relationship to reaching consensus. However, it has a significant negative relationship to the group decision quality. As was the pattern for giving and asking for opinions, the obverse, asking for orientation, also has a significant positive relationship to reaching group consensus, but a negative relationship to quality of decision.

The results for categories one and twelve contain some surprising findings. One would think that showing solidarity would be related to reaching consensus. It has a small negative relationship. However, it is significantly positively related to quality of group decision. In this category the CC groups had significantly more communication units. Another surprising finding demonstrated that showing tension is significantly positively related to reaching

consensus. One might think that it would hamper reaching consensus, but apparently it is better to get these tensions out than to fail to express them, in terms of the group's subsequent ability to reach consensus. However, both of these categories have such a small number of communication units that the apparent relationships may be a result of the fact that in the first case, showing solidarity occurs more in CC, which for other reasons has less solidarity, and the other communication processes described above working in favor of a higher quality solution. Likewise, the results for showing tension may be affected by its significantly greater occurrence in FtF.

The measure of quality of decision allows us to use the more powerful Pearson's coefficient of correlation, which the next table shows broken down for the FtF groups and the CC groups, as well as for all groups.

Figure 4-11

% of Group Communications in Each Bales Category, by Quality of Group Decision and Mode

Pearson's Correlations and Level of Significance

	all	ftf	cc
1. Shows solidarity	.683	.765	.535
2. Shows tension release, jokes	-.476	-.420	-.239
3. Agrees	.050	.363	.556
4. Gives suggestions	.444	.631	.-.022
5. Gives opinions	.119	.075	-.386
6. Gives orientation	-.408	-.580	-.074
7. Asks for orientation	-.276	-.067	-.107
8. Asks opinion	.285	-.780	.760
9. Asks for suggestion	.019	.265	-.008
10. Disagrees	-.196	.010	-.290
11. Shows tension	.122	.643	-.462
12. Shows antagonism	.134	.211	.049

Significance values for Pearson's R for 8 pairs of scores:

.10 - .549, .05 - .632, 02 - .685, .01 - .735

For 16 Pairs of scores (Pearson's R)

.10 - .400, .05 - .468, .02 - .542, .01 - .590

In most cases, the relationship between quality of the final group decision and percentage of interaction units in a category is in the same direction for CC and FtF. However, there are some exceptions that are notable. We are not sure how to interpret the differences. Agreement is strongly related to quality of decision in FtF, but not in CC. Giving orientation has a strong negative relationship for FtF, but only a very weak relationship for CC. Asking for opinions has a strong, significant NEGATIVE relationship for FtF, and a strong, significant positive relationship in CC. Showing tension has a strong positive relationship for FtF, and a negative relationship for CC. Thus, we see that the process is related to the outcome of the decision in different ways for the two media. They have their own unique dynamics, and what is effective in one medium may be counterproductive or ineffective in the other. An experiment designed to purposely manipulate these process variables might give us more insight.

In the next table, we see the correlations between Bales process categories and ability to reach consensus, by mode. A serious problem in looking at correlations for the face to face condition for this measure is that we are not dealing with much variance to explain... half the face to face groups, it will be remembered, were "tied " for top place with perfect 1.00 Kendall's coefficients, and the others were all above .98. Therefore, any apparent contrasts must be subjected to much further study, using a problem if possible which would not always result in complete consensus in face to face

groups. However, it appears that once mode is controlled, joking and laughing ("showing tension release") is not correlated to consensus, particularly for computerized conferencing.

Our final procedure in trying to trace differences in the relationship between process and outcome variables for the two modes of communication was to do a stepwise multiple regression. The first of these is shown in Figure 4-13. This analysis is of best predictors of quality of decision in the face to face groups.

The stepwise multiple regression proceeds by finding which Bales category is the best single predictor of variations in the dependent variable, in this case, quality of decision in the face to face groups. We see that "Asking for opinion" was the best single predictor, accounting for 60% of the variance. Then, when the proportion of statements in that category is held constant, the next best predictor for the face to face groups is category 6, giving orientation. Together, these two variables explain 87% of the variance and produce a multiple correlation coefficient of .93. Adding the next two steps produces statistically significant improvements in the prediction of quality of decision, though not large differences, since the first two predictors have accounted for most of the variance.

Figure 4-14 shows that the best predictors and combination of predictors is somewhat different for the CC groups. Asking for opinions is the most important predictor, accounting for 58% of the variance, just as it is the most important in the face to face

groups. Showing tension release also appears in the top four, as in face to face groups. However, unlike the face to face groups, asking for and giving suggestions are important predictors, and the agreement and giving orientation categories are not important.

The stepwise multiple regressions for amount of consensus are included for completeness' sake, though as we have noted above, there is so little variability in the face to face groups that the significance of these findings is problematic. The best two predictors for the FtF mode (giving suggestions and asking for orientation) are completely different than those for CC (giving opinion and showing tension release), but the third variable, asking for suggestions, is the same.

As with the simple correlations with mode, the dynamics of effective communication for the two media appear to be different, and are worthy of further investigation.

Figure 4-12

% of Group Communications in Each Bales Category, by Group Consensus, by Mode of Communication

Spearman's Rho Correlations and Level of Significance

Category	FtF	CC
	Spearman	Spearman
1. Shows solidarity	.457	-.095
2. Shows tension release, jokes	.051	-.524
3. Agrees	.342	.357
4. Gives suggestions	.837	-.214
5. Gives opinions	-.406	-.524
6. Gives orientation	.178	.405
7. Asks for orientation	.254	.595
8. Asks opinion	-.710	-.262
9. Asks for suggestion	.507	.755
10. Disagrees	-.292	.643
11. Shows tension	.228	.214
12. Shows antagonism	-.057	.476

Critical values for Spearman's Rho by Level of Significance

.10=.425
.05= .506
.01= .665

FIGURE 4-13

STEPWISE MULTIPLE REGRESSION

FACE TO FACE CONDITION

PERCENTAGE OF GROUP COMMUNICATION IN BALES CATEGORIES

BY QUALITY OF GROUP DECISION

STEP 1

VARIABLE SELECTED - CATEGORY 8 - ASKS FOR OPINION

Proportion of Variable Y Reduced	.608
Cumulative Proportion Reduced	.608
Multiple Correlation Coefficient	.780
F for Analysis of Variance (D.F. = 1,6)	9.294

STEP 2

VARIABLE SELECTED - CATEGORY 6 - GIVES ORIENTATION

Proportion of Variable Y Reduced	.267
Cumulative Proportion Reduced	.875
Multiple Correlation Coefficient	.935
F for Analysis of Variance (D.F. = 2,5)	17.508

STEP 3

VARIABLE SELECTED - CATEGORY 3 - AGREES

Proportion of Variable Y Reduced	.045
Cumulative Proportion Reduced	.920
Multiple Correlation Coefficient	.959
F for Analysis of Variance (D.F. = 3,4)	15.427

STEP 4

VARIABLE SELECTED - CATEGORY 11 - SHOW TENSION RELEASE

Proportion of Variable Y Reduced	.040
Cumulative Proportion Reduced	.961
Multiple Correlation Coefficient	.980
F for Analysis of Variance (D.F. = 4,3)	18.420

STEPWISE MULTIPLE REGRESSION

COMPUTER CONFERENCING CONDITION

PERCENTAGE OF GROUP COMMUNICATION IN BALES CATEGORIES

BY QUALITY OF GROUP DECISION

STEP 1

VARIABLE SELECTED - CATEGORY 8 - ASKS FOR OPINION

Proportion of Variable Y Reduced	.577
Cumulative Proportion Reduced	.577
Multiple Correlation Coefficient	.760
F for Analysis of Variance (D.F. = 1,6)	8.190

STEP 2

VARIABLE SELECTED - CATEGORY 2 - SHOWS TENSION RELEASE

Proportion of Variable Y Reduced	.142
Cumulative Proportion Reduced	.719
Multiple Correlation Coefficent	.848
F for Analysis of Variance (D.F. = 2,5)	6.389

STEP 3

VARIABLE SELECTED - CATEGORY 9 - ASKS FOR SUGGESTION

Proportion of Variable Y Reduced	.129
Cumulative Proportion Reduced	.848
Multiple Correl tion Coefficient	.921
F for Analysis of Variance (D.F. = 3,4)	7.430

STEP 4

VARIABLE SELECTED - CATEGORY 4 - GIVES SUGGESTION

Proportion of Variable Y Reduced	.123
Cumulative Proportion Reduced	.971
Multiple Correlation Coefficient	.985
F for Analysis of Variance (D.F. = 4,3)	25.103

STEPWISE MULTIPLE REGRESSION

FACE TO FACE CONDITION

PERCENTAGE OF GROUP COMMUNICATION IN BALES CATEGORIES

BY KENDALL'S COEFFICIENT OF CONCORDANCE

STEP 1

VARIABLE SELECTED - CATEGORY 4 - GIVES SUGGESTION

Proportion of Variable Y Reduced	.408
Cumulative Proportion Reduced	.408
Multiple Correlation Coefficient	.639
F for Analysis of Variance (D.F. = 1,6)	4.143

STEP 2

VARIABLE SELECTED - CATEGORY 7 - ASKS FOR ORIENTATION

Proportion of Variable Y Reduced	.350
Cumulative Proportion Reduced	.759
Multiple Correlation Coefficient	.871
F for Analysis of Variance (D.F. = 2,5)	7.867

STEP 3

VARIABLE SELECTED - CATEGORY 9 - ASK FOR SUGGESTION

Proportion of Variable Y Reduced	.145
Cumulative Proportion Reduced	.904
Multiple Correlation Coefficient	.951
F for Analysis of Variance (D.F. = 3,4)	12.487

STEPWISE MULTIPLE REGRESSION

COMPUTER CONFERENCING CONDITION

PERCENTAGE OF GROUP COMMUNICATION IN BALES CATEGORIES

BY KENDALL'S COEFFICIENT OF CONCORDANCE

STEP 1

VARIABLE SELECTED - CATEGORY 5 - GIVES OPINION

Proportion of Variable Y Reduced	.550
Cumulative Proportion Reduced	.550
Multiple Correlation Coefficient	.741
F for Analysis of Variance (D.F. = 1,6)	7.328

STEP 2

VARIABLE SELECTED - CATEGORY 2 - SHOWS TENSION RELEASE

Proportion of Variable Y Reduced	.319
Cumulative Proportion Reduced	.869
Multiple Correlation Coefficient	.932
F for Analysis of Variance (D.F. = 2,5)	16.573

STEP 3

VARIABLE SELECTED - CATEGORY 9 - ASKS FOR SUGGESTION

Proportion of Variable Y Reduced	.112
Cumulative Proportion Reduced	.981
Multiple Correlation Coefficient	.990
F for Analysis of Variance (D.F = 3,4)	67.920

Summary

We compared the explicit, verbalized content of communications in a face to face conference with those in a computerized conference, using Bale's original (1950) categories for Interaction Profile Analysis. The observed differences between the communications modes are:

1) There is significantly more "showing solidarity" in CC.

2) There is more "tension release" (joking, laughing), agreement, and disagreement expressed in face to face groups.

3) Asking for and giving opinions and giving suggestions occur more mn CC.

4) Asking for information or clarification occurs more in FtF.

These differences in interaction process are somewhat task depmndent and are related to differences in outcome of the meeting in somewhat different ways for the two communication modes.

For both modes, quality of decision is positively related to the proportion of communications showing solidarity and agreeing; and negatively related to showing tension release and giving orientation. However, asking for opinions is negatively related to quality for FtF and positively for CC. The opposite is true for showing tension; it

is positively related for FtF and negatively for CC. For both modes, the proportion of statements asking for opinions is the best predictor of quality of decision, followed by the proportion showing tension release. The percentage of communications in these two categories account for 93% of the variance in the quality of decision in the FtF groups, and 85% in CC.

We thus have some insight into the apparent puzzle that there is something mo2e efficient per communication unit in CC. With only half the communication units in the same amount of elapsed time, the CC groups reached the same improvement in quality of solution. This seems to be accounted for by the greater proportion of asking for opinions and the lesser proportion of tension release in CC.

Differences in ability to reach consensus must be interpreted with caution since there was so little variability in the FtF condition. With this caveat, we found that agreement is positively related to consensus in both modes (as would be expected), and so is giving suggestions. Giving opinions is negatively related to consensus in both modes. However, giving suggestions is positively related for FtF, but not for CC, whereas disagreement is positively related for CC and not for FtF, and showing tension release is negatively related for CC but not for FtF.

The stepwise multiple regressions to identify the best predictors of consensus for the two modes give coMpletely different results. In FtF, giving suggestions and asking for orientation are the most powerful predictors. For CC, giving opinions and showing tension

release are the most powerful predictors (the less, the better).

The findings are intriguing and suggest that further investigation would be fruitful. Among the variations which would help to establish the extent of generalizability of our findings are

1) Other forms of CC, including more structured conferences and asynchronous, longer term conferences with more experienced users.

2) A wider variety of tasks, including one that does not generate complete consensus in most FtF groups.

3) Isolation and examination of the role of non-verbal communication in FtF, and how this is substituted for in CC.

Chapter 5

Processes of Decision Making II: Inequality of Participation and Dominance

Whereas the experimental work on small group behavior in face to face meetings shows a tendency towards inequality of participation and dominance by a single member, this has not been observed to be true in field trials of computerized conferencing. For example, observations of behavior on FORUM have led to the conclusion that "greater equality in group participation can be facilitated by the use of computer conferencing, especially in synchronous sessions". (Ferguson and Johansen, 1975 and Vallee, Johansen, Lipinski, Spangler and Wilson, 1975, summarized in Johansen, Vallee, and Spangler, 1979, p. 151).

There is a tendency towards inequality of participation more often in face to face groups than in computerized conferencing groups. This seems to be related to the lack of leadership/dominance in CC, and the consequently greater difficulty in achieving consensus.

We actually have two different phenomena here which can be measured, related to inequality. The first has to do with equality of participation among all members of a group. This is measured by an index of inequality, which can be computed on number of turns or on number of participation units, measured in Bales IPA units. Though fairly equal participation does tend to be somewhat higher in computerized conferencing, the differences are not statistically

significant.

The second measure is of dominance or leadership. This focuses only on the proportion of the interaction accounted for by the most active individual. We arbitrarily chose the cutoff point of one individual in the five person group contributing a third or more of the discussion to indicate dominance by that individual. When dominance is measured in this manner, face to face groups are significantly more likely to generate a dominant person or "leader" in the unstructured, value laden "Forest Ranger" problem.

Measure of Inequality of Participation

An index of inequality of participation in a group was generated using the same approach as economists use in constructing a Lorenz curve to get a coefficient which will describe inequality of distribution of income in a society. It compares the cumulative percentage of statements made, starting with the least active participant, against the cumulative percentage of the number of participants. This index is constructed in such a way that it yields a value of 0 if there is total equality of participation, and 1 if there is total inequality, regardless of the size of the group. The numerator represents the observed differences between the proportions of statements made by each of the participants and the proportions they would have made if each contributed an exactly equal share. The denominator consists of the maximum value which this sum of observed differences could possibly reach in a group that size in which there was total inequality, with one of the members making all of the statements. Thus, the index compares observed inequality to the maximum possible for a group that size, according to the following formula:

Let I = Index of inequality

N = Number of members in group

Oi= Observed cumulative proportion of statements

Ei= Expected cumulative proportion if there were total equality of participation; equal to cumulative proportion of total number of members of group.

I=(1/N Sum (Ei-Oi))/1/2 (1-1/n)

This index was first computed on number of turns.

We see that in 10 of the 16 groups, the index of inequality was higher in the face to face condition, for the same group. Thus, there is some tendency for face to face discussions to have more unequal participation. The T of 50 on the Wilcoxan matched pairs test shows that the differences are not statistically significant.

However, the very largest indices are for some CC groups. Looking at the transcripts, we discovered that in those groups, one or two individuals participated very little-- they entered one or two comments, and then seemed to become confused and/or passive, and were unable to keep up with the discussion. These tended to be older individuals. Their entries also tended to be very long, because they kept forgetting how to enter a comment, so they would have many, many lines in a single "turn" or comment when it was finally entered.

Figure 5-1

INDEX OF INEQUALITY OF PARTICIPATION (NUMBER OF TURNS), by COMMUNICATION MODE

group	face to face	cc
1a	.312	.103
1b	.204	.133
1c	.354	.071
1d	.265	.136
2a	.098	.118
2b	.105	.122
2c	.384	.167
2d	.411	.216
3a	.157	.125
3b	.198	.388
3c	.189	.132
3d	.197	.134
4a	.160	.30
4b	.143	.133
4c	.156	.537
4d	.288	.383

Wilcoxan matched pairs test: T=50, n=16, p>.05.

A second method of analysis, which takes problem as well as mode into account, is the analysis of variance of the sixteen indices. The indices in these tables were computed on number of Bales IPA units, rather than number of turns. This method of analysis also indicates no statistically significant differences between modes in the overall equality of participation. We do note, however, that the average inequality for face to face groups discussing the Forest Ranger problem (.33) is strikingly larger than for the other problem/mode combinations.

Most of the indices are quite low. In other words, for some reason we had fairly equal participation in both the face to face and computerized conferences in this experiment. This led us to the speculation that perhaps something related to the experimental sequence or treatment was producing the equal participation pattern, which is not usual for human groups. We think that one reason why there may not be any significant difference between the communication modes in terms of equality of participation in this experiment is that all groups were trained on the computer before they were actually run in groups. In the first experiment (pilot), this was not true. If indeed the terminals lead to greater equality of participation, then the tendency of everyone in the group to add comments may have already been set in the pretraining session. Thus, in FtF conditions, people who normally would be hesitant to speak in a strange group may have been more at ease, due to their common experience in the training.

However, another problem is in our initial choice of measure. The index does indeed measure how close to equal participation all members of the group are. However, a related but different question is the extent to which a single individual is able to dominate the interaction. Suppose, for instance, that four members of the group contributed 16% each of the units, and the fifth, the remaining 36%. Our index would not be particularly high for average inequality, because four of the five members are very close to the expected equal participation rate of 20%. Our index does not pick up the emergence of a single dominant individual in a leadership position, and this is one of the objectives of the experiments-- to see if there is any difference between CC and FtF in the tendency for a dominant individual to emerge.

Therefore, we devised a more primitive way of checking for this. We think that a rough indicator of a leader in the five person group is that one person emerges with over 33% of the interaction units. This corresponds with Shaw's (1976, p.157) graphing of the original Bales study data (Bales, Strodtbeck, Mills, and Roseborough, 1951) to indicate that on the average, in five person groups, only one person had above 20% of the interaction units. "Over a third" of the interaction was simply picked as a figure that would undoubtedly indicate dominance in a five person group. Another more generalizable breaking point might be to set more than 50% above expected or equal participation to show a dominant rate of participation. This would have set a cutting point of 30% for this experiment. However, we suspect that for very large groups, there is some absolute minimum proportion of the interaction necessary in

order to create leadership/dominance, and that some adjustment factor would have to be added.

Figure 5-2
Inequality Measures

Arctic

Means for Index of Inequality, Bales IPA Units
Mode of Communication

		FTF	CC	
Order of Problem	1st	.16705	.169675	.168362
	2nd	.27415	.196025	.235087
		.2206	.18285	

2x2 CRANOV *

Source	SS	df	MS	F
A	.005699	1	.005699	1.029939
B	.017809	1	.017809	3.218493
AxB	.006522	1	.006522	1.178674
Wg	.0662	12	.005533	
Total	.096428	15		

Table Value for F
1 and 12 df = 4.75
Not significant

A = mode
B = order
WG = error term

* CRANOV stands for a Completely Randomized Analysis of Variance design.

Figure 5-3
Inequality Measures

Forest Ranger

Means for Index of Inequality, Bales IPA Unids
Mode of Communication

		FTF	CC	
Order of Problem	1st	.3193	.25075	.285025
	2nd	.334825	.2229	.278862
		.327062	.236825	

2x2 CRANOV

Source	SS	df	MS	F
A	.032571	1	.032571	2.565049
B	.000152	1	.00152	.01197
A x B	.001881	1	.001881	.148133
WG	.52377	12	.012698	
Total	.186982	15		

Table Value for F
1 and 12 df = 4.75
Not significant

A = mode
B = order
WG = error term

Though we did not expect it before the experiment, we were led by our findings in figures 5-2 and 5-3 to look separately at dominance for dhe forest ranger problem, which appeared to show much more dominance and inequality than the other problem.

The following table gives the number of people by mode who used the indicated percentage of Bales units in the discussion for the Forest Ranger problem.

Figure 5-4

Dominance in the Forest Ranger Problem, by Mode

% Range	CC	FTF
0 to <5		1
5 to <10	3	4
10 to <15	8	9
15 to <20	7	6
20 to <25	8	10
25 to <30	11	4
30 to <35	2	1
35 to <40	1	3
40 to <45		1
45 to <50		
50 and over		1

Let us now pull out just the dominant individual from each group.

Largest %	FTF	CC
33%+	5	1
<33%	3	7

Chi Square=4.16, p=<.05

In five of the eight face to face groups, a single individual dominated the discussion, contributing over a third of the communication units. In only one of the eight computerized conference groups did such a dominant individual emerge.

The chi square test is not fully appropriate with this small a number of cases, but the expected number of cases per cell is close enough to five to enable it to serve as a rough test of significance.

When an analysis of variance is performed, the fact that we have only sixteen observations also makes it difficult to reach high levels of statistical significance. The table for the analysis of variance follows, however. It shows that the differences in dominance reached something between the .05 and .10 level of significance.

Figure 5-5
Dominance by Mode of Communication, Forest Ranger Problem
2 x 2 CRANOV (16 observations)
Maximum % Participation for the Most Prolific Member

Means

		Mode of Communication FtF	CC
Order of Problem	1st	35.12	30.25
	2nd	36.85	27.53
	Average	35.98	28.89

CRANOV

Source	SS	df	MS	F
A	201.18	1	201.18	3.618*
B	.98	1	.98	.002
A x B	19.78	1	19.78	.356
WG	667.31	12	55.61	
Total	889.25	15		

Table value for F
p =.10 1 and 12 df=3.18
* Significant

A = mode
B = order
WG = pooled error term

The Cranov design yields a significant difference between the FtF and CC conditions. The maximum percentage of participation for the most prolific member is greater in the FtF conditions.

This is not the pattern demonstrated for the scientific problem, Arctic, as shown below.

Figure 5-6

Arctic Problem, Distribution of % of Bales IPA Units Contributed

% Range	CC	FtF
0 to <5		2
5 to <10		
10 to <15	8	7
15 to <20	11	8
20 to <25	12	13
25 to <30	7	8
30 to <35	1	2
35 to <40	1	
40 to <45		
45 to <50		
50 and over		

Both media of communication are shown to be lacking the emergence of a single dominant person, in most groups, for this problem.

It will be remembered that it was the Forest Ranger problem for which there was the tremendous difference between face to face groups and CC groups in ability to reach total consensus. It seems plausible that this is strongly related to the much greater tendency for a single dominant leader to emerge on this value-laden kind of problem

in the face to face meeting.

However, there is probably a stronger factor at work here than difference in the type of problem. The Arctic problem used a set of procedures and instructions that probably created a communications structure that is conducive to equal participation of group members and not conducive to the early emergence of a single dominant leader due to the "latency of verbal response" phenomenon (Willard and Strotbeck, 1972).

The Structure of the Communication Process for Arctic

For the Arctic problem, even in the face to face condition, each individual first read the problem alone (as with Forest Ranger) and then INDEPENDENTLY ARRIVED at an initial solution, WROTE IT DOWN, and brought his/her independently generated solution to the room to begin the face to face discussion.

This corresponds to the first stage of a "brainstorming" technique (Osborn, 1957) for structuring face to face meetings, and also has similarities to stage one of the "Nominal group technique" (Van de Ven and Delbecq, 1974).

One study of the effect of such structuring was done by Vroom, Grant, and Cotton (1969). Among the communication structures they contrasted were those in which

1) Members interacted with one another during the generation of solutions, but were prevented from interacting during the evaluation

of solutions.

2) members were prevented from interacting with one another during the generation of solutions, but did interact during the evalution of solutions.

3)members interacted with one another during both generation and evaluation of solutions.

4)members were prevented from interacting with one another during both generation and evalution of solutions (Vroom, Grant, and Cotton, 1969, p.77).

What we did in Arctic was to create a structure such as (2) above, in which the group members did not interact during initial solution generation, but did interact during evaluation of the solutions. Each came with his/her own written solution as the basis for starting the face to face discussion; and almost all groups, at the beginning of their discussion, began with each individual presenting his/her solution. In the CC condition, this was done by a volunteering of "sharing of the rankings". We created a special command, "+share rank", which produced a table of their fifteen ranks and entered it as a conference comment. In the face to face condition, it was usually done by each person reciting their topmost set of items, and/or passing around their ranking sheets (we did not give them a blackboard).

Vroom et. al. found that groups in which members were prevented from

interacting during the solution-generation phase produced a larger number of different solutions, and more high-quality solutions. The study did not include measures of equality of participation in discussions, however.

Another related study by Carlston (1977) looked at the effect of polling order on social influence in decision making groups. Their results indicate that "speaking order necessarily mediates the effects of social influence processes in discussion groups"(p. 122). The independent variable in this study was whether groups were left to "voluntarily" determine initial speaking order, or to follow a speaking order predetermined by the experimenter. The dependent variable was conformity; and they found that the probability of any subsequent speaker moderating his/her opinion towards that of the members who had already spoken did increase (The overall probability of conformity was 47.5% for second speakers, 62.5% for third speakers, and 77.6% for fourth speakers, in these four person groups) (Carlston, 1977, pp. 119-120). In the study, the subjects had independently recorded their opinions on a pre-discussion questionnare, but did not bring their written opinions with them to the face to face discussion.

Dominance and Quality of Decision for Forest Ranger Problem

An interesting correlation exists between the proportion of interaction units contributed by the most active group member (dominance) and the decision made about whether Bill's pay should be reduced or maintained. This is the decision atom for which there was unanimous agreement by our three expert judges that the better

decision was to maintain Bill's pay, rather than to punish him by reducing it. Figure 5-7 shows that within the media, the proportion of interactions accounted for by the most dominant individual is higher for groups that decided to punish Billy by reducing his pay.

Figure 5-7

Dominance by Decision for Forest Ranger

Mode	Reduce Bill's Pay	Maintain Pay
CC	28,28,30,31,36	25,26,27
Majority		
FtF	30,36,37,37,44,53	23,29

A Mann-Whitney U test on the relationship between the proportion of comments contributed by the most active individual (regardless of medium) and the decision made with regard to lowering or maintaining Bill's pay showed the relationship to be significant at the .002 level.

This correlation suggests the possibility of a threshold of dominance level beyond which the group is likely to make a more punitive or extreme decision than the individuals might otherwise be likely to support.

Age and Reactions to CC

In this first experiment, we foolishly relied upon the assistants to code an approximate age for the subjects. The reasoning was that we did not want to sensitize the subjects to age differences. The problem with this approach, discovered too late, is that for 15 of the 80 cases, all of whom were above-college-age subjects, the

assistants did not want to guess if the subject was over or under 40, and so recorded nothing.

Another problem, discovered too late, was that the pilot tests were almost all done on 21 and under age students. They had no problem learning the use of CC in approximately 20 minutes. About a third of the way through the experiment, we realized that many of our older subjects needed longer than 20 minutes to become comfortable with the medium. In experiment two, we plan to give a full hour's training and practice. This will be longer than most subjects need, but will better assure that older subjects have sufficient time to learn to use the computer terminal and the commands taught, so they are not at a disadvantage in the group discussion.

Thus, the data in Table 5-8 should be taken as suggestive of a difference, and not definitive. It shows quantitatively one aspect of the correlation which we observed between age and ease of adaptation to CC. One sees that almost two thirds of the older subjects took very little part in the computerized conference, entering 5 or fewer comments over 60 to 90 minutes (both problems, Forest Ranger and Arctic, are combined for these data). There were exceptions, of course; some of the older subjects were among the fastest learners and most active participants. But our general conclusion is that there is a tendency for older persons who have never used a computer terminal to take somewhat longer to become adroit at using the medium, and that in the future, training procedures and time should be adapted to make sure that older persons are not put at a disadvantage by training procedures that are

inadvertently geared towards younger persons.

Figure 5-8

CC: Number of Turns, by Age

Age	5 or less	6-10	11-19	20 or more	N
21 or under	10%	16	57	16	49
22-39	0	62%	38	0	8
40 or over	62%	13	0	25	8

Summary

We have seen that there is some tendency for greater equality of participation in computerized conferencing than in face to face discussions, but that there are many exceptions-- so many that the difference is not statistically significant.

Looking at dominance or leadership by a single individual, we observed that in an unstructured discussion of a value-laden human relations problem, a dominant individual was able to emerge in a face to face discussion, but not in a computerized conference. We have noted that the presence of such a dominant person seemed to permit the face to face groups to reach consensus on the decision, whereas CC groups could not.

On the other hand, structuring of a communications process, both in face to face and computerized conferencing, can effectively be done to assure that all individuals have the opportunity to be equally heard. (Of course it is a lot easier to effectively create such special structures for communications mediated by the computer.) When such structured communications rules were introduced, the face to face mode as well as the CC mode did not tend to permit a dominant individual to emerge as the leader of a discussion. There is an alternative explanation, however. Perhaps the nature of the Arctic problem as a knowledge pooling task encourages more equality.

Whether one wishes to have a dominant individual emerge to lead group decisions depends upon one's objectives. It does aid the emergence of total group consensus relatively quickly. However, it may also lead to agreement on a poor decision, advocated by the dominant individual. Those face to face groups that had dominant members tended to agree on a decision on the Forest Ranger problem that was judged to be very poor by the experts.

There is some indication that persons over 40 have difficulty adjusting to CC and may not make very many comments in discussions in this medium. However, we had a small number of such subjects, so the results can only be taken as suggestive of something deserving further study. In addition, they were in a minority position, which might also account for their relatively lower average participation.

Chapter Six

Subjective Satisfaction

Following the two group discussions on two media, all subjects were asked to complete a pair of questionnaires, one for their face to face discussion, and an identical one for their computerized conferencing discussion. They were explicitly told that they would be answering the same questions for the two discussions, in order for us to be able to compare their reactions.

A correlated T test for paired comparisons was used to test for significant differences in the responses to the questions between the face to face and the computerized conferencing condition. In each of these T-tests, there were 80 responses (paired) and 79 degrees of freedom. In Figure 6-1 are the questions asked and the detailed results of the T-test for statistical differences. We adopted the .05 level of significance, and will consider any difference which has a higher than 5% probability of occurring by chance to be "not significant".

We note that in computerized conferencing, the issues seemed clearer. There were no significant differences in overall pleasantness of the experience or satisfaction with one's performance. The subjects did perceive the significantly greater difficulty in reaching consensus via CC.

Questions 9 through 17 on the post-experimental questionnaire were

scales originally devised by the Communications Studies Group in Great Britain. They are called the " "DACOM" (Description and Classification of Meetings) scales, and have been used in many other studies. We see that for new users of computerized conferencing, the medium seems satisfactory for most communication purposes, but significantly less satisfactory than face to face meetings. The next part of this paper shows that these perceptions appear to change as more experience is gained with the medium.

Figure 6-1

T Tests for Subjective Reactions to Communications Media

6. The problem was:

: 1 :	2 :	3 :	4 :	5 :	6 :	7 :
Completely interesting			Neutral			Completely Boring

Mean CC	Mean FtF	T	Prob
2.49	2.62	-0.72	.48

7. The situation struck me as:

: 1 :	2 :	3 :	4 :	5 :	6 :	7 :
Realistic						Unrealistic

Mean CC	Mean FtF	T	Prob
2.52	2.66	-.86	.39

8. The issues involved were:

: 1 :	2 :	3 :	4 :	5 :	6 :	7 :
Completely Clear						Completely Unclear

Mean CC	Mean FtF	T	Prob
2.31	2.74	-2.18	.03

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.

9. Giving or receiving information

Mean CC	Mean FtF	T	Prob
3.55	1.79	7.59	.00

Figure 6-1, cont.

10. Problem solving

Mean CC	Mean FtF	T	prob
4.42	1.98	10.57	.00

11. Bargaining

Mean CC	Mean FtF	T	prob
4.41	2.13	9.23	.00

12. Generating ideas

Mean CC	Mean FtF	T	prob
3.06	1.64	6.73	.00

13. Persuasion

Mean CC	Mean FtF	T	prob
4.10	2.12	8.81	.00

14. Resolving disagreements

Mean CC	Mean FtF	T	prob
4.46	2.40	8.71	.00

15. Getting to know someone

Mean CC	Mean FtF	T	prob
3.94	2.28	6.26	.00

16. Giving or receiving orders

Mean CC	Mean FtF	T	prob
3.23	3.08	.56	.58

17. Exchanging opinions

Mean CC	Mean FtF	T	prob
3.45	1.59	8.26	.00

Figure 6-1, cont.

The following questions deal with your feelings about your group and its discussion and your participation today.
Once again, we ask you for a rating of between 1 (top rating) and 7 (bottom rating)

18. Taking part in this research was

: 1 : 2 : 3 : 4 : 5 : 6 : 7 :
Pleasant Neutral Unpleasant

Mean CC	Mean FtF	T	prob
1.74	1.59	1.18	.24

group discussion?

: 1 : 2 : 3 : 4 : 5 : 6 : 7 :
Completely Satisfied Completely Unsatisfied

Mean CC	Mean FtF	T	prob
2.60	2.26	1.72	.09

20. Did your group reach a consensus?

: 1 : 2 : 3 : 4 : 5 : 6 : 7 :
Definitely Yes Not at all

Mean CC	Mean FtF	T	prob
3.74	1.46	10.58	.00

Comparative Data From Long Term Field Trials and Other Experiments

The Electronic Information Exchange System was designed to enhance communication within geographically dispersed "small research communities", "conceived as groups of 10 to 50 individuals sharing an interest in a scientific or technological problem area." (NSF 76-45, p.3) The Division of Science Information (now the Division of Information Science and Technology) of the National Science Foundation issued a program announcement in 1976 inviting proposals for "operational trials" of the system. Four groups were initially chosen to participate, beginning in the fall of 1977, and three more started subsequently.

The Division of Mathematical and Computer Research funded a study by Hiltz to conduct an across-groups assessment of the impact of the use of EIES, which included a series of questionnaires before use, at approximately three months after use began, and at approximately 18 months. The three months follow-up included the same CSG subjective satisfaction scales as were used in the controlled experiments.

The perceptions of individuals about how useful and satisfactory this system is appear to change markedly with experience. The data in figure 6-2 can give us a rough idea of the extent to which perceptions of computerized conferencing as a medium of communication change with experience. We can also get an idea of how our experimental results for EIES compare to audio and video conferencing and another computerized conferencing system, Planet.

Planet is a simple conferencing system, comparable to the limited functions and limited number of commands taught to the subjects in the controlled experiment using EIES. The audio conference syqtem was the "remote meeting table", which has a name plate and a light that lights up at the place "saved" for the microphone of each of the participants at other sites, whenever that person is speaking. The video conference has T.V. screens at each of two locations, which show the top portion of the attendees at the other conference site, as well as carrying an audio channel.

The data are NOT directly comparable; the only thing the various groups have in common is that they were asked the same questions, the CSG "DACOM" scales, following the use of a communications medium. The question asked was, "How satisfactory do you think this medium would be for the following kinds of activities or processes?" Respondents were then given a series of one to seven scales that ranged from completely satisfactory to completely unsatisfactory.

The subjects and the task varied widely, from a completely structured laboratory experiment to totally unstructured field trials. Thus, it would not be warranted to use tests of statistical significance to compare the differences in scores. Where there are differences, they may be due to these other sources of variation, rather than only to the difference in medium of communication.

Having stated these limitations in interpreting the data, what are the most interesting things in the table? First of all, the communication task categories have been arranged from those for which

computerized conferencing seems to be relatively good, compared to face to face, to those for which it is relatively poor. Face to face is listed first, because this is usually taken as the "standard of comparison". Note that people do NOT consider face to face meetings "completely satisfactory", particularly for a routine managerial task such as "giving and receiving orders".

For the pure "information exchange" tasks, experienced users of EIES find it as good as or better than face to face communication.

It is in the areas of ACTING on information and reaching a decision (bargaining, resolving disagreements, persuasion) that computerized conferencing is seen as clearly not as satisfactory as a face to face meeting. However, it is still rated on the "satisfactory" end of the scales.

Given this relative area of weakness, the focus of our next series of controlled experiments will be on attempting to create "decision aid" tools that may enable a group to bargain, persuade, and resolve disagreements more effectively than in an unstructured computerized conference.

Secondly, we notice that there is a significant increase in ratings of EIES as a function of time on line. Of course, some of this may be self selection; those who do not find it satisfactory never use it enough to become "experienced". It seems to take considerable time before people feel completely comfortable and skilled at using this new medium of communication. The same is probably also true for face

to face meetings and the telephone, but it is not as obvious, since this learning and acclimation have taken place at an earlier point in the lives of participants.

The ratings of computerized conferencing seem to be fairly generalizable across systems. The Planet ratings are by respondents whose hours on line span those for the new, intermediate, and experienced EIES users, and most of the PLANET ratings do lie within the range spanned by the EIES scores. The exception to this is "Getting to know someone". This is probably due to a design difference. Planet does not have a directory where one may read descriptions of all of the members of the system, and pick out someone with similar interests with whom to communicate. Nor does it encourage the sending of private messages among subgroups-- very important in clique-building. Finally, PLANET does not have the ubiquitously on-line "user consultants" who advise newcomers on people with whom they might like to communicate, as well as on the mechanics of system usage.

The comparisons to audio and video, which are often considered more "wideband" or "natural" forms of communication, may be surprising to some. For those tasks for which comparable ratings were reported, computerized conferencing is rated at approximately the same or higher level of adequacy.

Figure 6-2: Comparison of Media

TASK	FtF	Exp EIES	INT EIES	New EIES	PLANET	VIDEO	AUDIO
Giving and Receiving information	1.8	1.8	2.2	3.6	2.3	2.2	2.6
Giving or receiving orders	3.1	2.3	3.3	3.2	2.8		
Exchanging opinions	1.6	1.7	2.2	3.5	2.5	1.9	
Generating ideas	1.5	2.4	2.8	3.1	2.8	2.7	
Getting to know someone	2.3	3.2	3.8	3.9	4.8	4.0	5.1
Problem Solving	2.0	2.7	3.7	4.4	3.9	2.7	
Bargaining	2.1	3.1	4.0	4.4	4.3	3.6	3.9
Persuasion	2.1	3.4	4.1	4.1	4.4	3.6	3.9
Resolving Disagreements	2.4	3.3	4.3	4.5	4.1		

KEYS

FtF	Face to face discussion, experimental subjects,N=80.
Exp Eies	Experienced EIES users in the operational trials with fifty or more hours of experience on line. Follow-up Questionnaire at 3-6 months, N=19.
Int EIES	Intermediate EIES Operational Trials users with 5-49 hours on line. Follow ups at 3-6 months.N=76.
New EIES	EIES users with less than three hours on line. Experimental subjects answering the post-use questionnaire. N=80.
PLANET	Post-use questionnaires completed by 57 PLANET users. Source: Johansen, DeGrasse and Wilson, 1978. Scale reversal computed for comparability.
VIDEO	Confravision. Source, Champness, 1973a, reported in Pye and Williams, 1977.
AUDIO	"Remote Meeting Table", Champness, 1973b, reported in Pye and Williams, 1977.

Comparison to Westgate's Results

The only other published experiments which have compared FtF and CC were carried out by Westgate in 1977. He used 32 students from the MBA program at the Cranfield School of Management in Great Britain. Groups of three to four first played "Crisis", a competitive negotiation excercise, in either the face to face or CC mode, and then repeated the game with a different subgroup in the other mode. The CC systems used were FORUM, Confer, and ZCONFER. Training procedures were not specified in the report (Westgate, 1978). Thus, the group size, task, specific CC system and other experimental procedures all differed from the study reported here. Problems with frequent disconnects were reported (p. 20) and this can be expected to severely affect subjective satisfaction with the CC mode.

Among Westgate's dependent variables were the DACOM scales reported in this chapter. The means and standard deviations which he obtained are as follows:

Figure 6-3

DACOM Scales: Comparisons to Westgate's Results

Category	Westgate	EIES EXP
Exchanging Information	2.4	3.6
Giving or receiving orders	2.5	3.2
Exchanging opinions	3.6	3.5
Problem solving	4.3	4.4

Generating ideas	4.4	3.1
Bargaining	4.7	4.4
Resolving Disagreements	5.3	4.5
Getting to know someone	5.8	3.9

Thus, we have the somewhat puzzling finding that Westgate's subjects' ratings were higher for exchanging information and giving and receiving orders; and lower for all of the other functions. At least the two studies have similar results in one respect: the functions for which the new users of CC gave it the highest ratings were exchanging information, opinions and orders. As Westgate's factor analysis of the items points out, these can be considered "impressional" communications functions. However, the highest rated function among the EIES subjects, generating ideas, is not highly rated in the British experiment.

Whether these differences can be attributed to differences in task type, group size, specific CC system used, the disconnects reported for the British study, or other differences in experimental procedures cannot be determined.

Amount of Participation vs. Subjective Satisfaction

To what extent is the subjective evaluation of the experience of taking part in a group discussion via computer a function of one's facility in taking an active part in the written discussion? We attempted to measure this by cross-tabulating amount of participation (as measured by number of turns) with the post-experiment questionnaire items on subjective satisfaction with the discussion.

Our problem in this analysis is lack of variation in the dependent variables. The four subjective satisfaction scales are highly skewed towards the positive end, with practically all subjects checking point 1 (the highest) or point two on the seven point semantic differential scales. Thus, for example, in response to "Taking part in this discussion was .. Pleasant..... Unpleasant, 42 of 80 checked 1, and 27 checked 2, 5 checked 3, and only 6 checked 4 (neutral) or lower.

Participation as measured by number of turns is not entirely valid, since some of these turns are much longer than others. For this analysis, number of turns (number of comments entered into the computerized conferencing transcript, as counted by an analytic routine), was broken into five categories, ranging from "very low" to "very high". The analysis was repeated, using number of lines composed, in order to measure participation by total amount rather than number of turns. The amount of participation as measured by lines composed was also broken into five categories.

Besides pleasantness of taking part, the scales asked about the perceived friendliness of the group, perceived productivity of the discussion, and satisfaction with one's own performance in the discussion.

The only significant relationship was between number of turns and satisfaction with one's own performance (chi square =41.19 with 24 degrees of freedom; p=.015). However, even this relationship was weak (gamma= -.17).

Thus, we cannot find much of a relationship between amount of participation and subjective satisfaction with computerized conferencing. However, we think that this is because the measurement scales used in this study were too insensitive to variations. It will be remembered that there was a relatively high degree of equality of participation in this experiment. With little variability in the independent and dependent measures used, one can hardly expect to find a statistically significant relationship. We believe that such a relationship is only likely to appear in a longer term use of computerized conferencing, rather than in the synchronous 90 minute discussion used in this study.

Gender and Subjective Satisfaction

Cross tabulations were made of the individual items on the CSG scale (questions 9 to 17) by gender, with mode of communication controlled. The chi square tests showed no significant differences between males

and females in the CC mode.

Figure 6-4

T Test for Differences Between Problems

Computerized Conferencing Mode

Task	Mean for Forest Ranger	Mean for Arctic	T	p
Giving and Receiving Information	3.2	3.9	1.59	.116
Problem Solving	4.0	4.8	2.06	.042
Bargaining	4.1	4.6	1.52	.133
Generating Ideas	2.9	3.2	0.84	.403
Persuasion	4.1	4.1	0.0	1.000
Resolving Disagreements	4.6	4.3	1.00	.320
Getting$to Know Someone	4.0	3.8	0.55	.581
Giving or Receiving Orders	3.4	3.0	0.91	.365
Exchanging Opinions	3.2	3.8	1.49	.140

Figure 6-5

T Test for Differences Between Problems

Face to Face Mode

Task	Mean for Forest Ranger	Mean for Arctic	T	p
Giving and Receiving Information	1.6	1.9	1.41	.012
Problem Solving	1.8	2.1	1.14	.257
Bargaining	2.0	2.2	.52	.608
Generating Ideas	1.7	1.6	.57	.573
Persuasion	2.1	2.2	.19	.406
Resolving Disagreements	2.3	2.5	1.38	.319
Getting to Know Someone	2.2	2.4	.43	.668
Giving or Receiving Orders	3.0	3.1	.20	.884
Exchanging Opinions	1.6	1.6		

Subjective Satisfaction: Differences Between Problems

When mode of communication is controlled, there are few significant differences in the ratings for our two task types, within mode. For computerized conferencing, the only significant difference shown in Table xx is for problem solving, for which Arctic receives a somewhat lower rating. In the face to face mode, the only significant difference is for giving and receiving information, where once again the rating is lower for the more difficult problem, Arctic. In both cases the differences are small, even though statistically significant. Thus, we come to the conclusion that the ratings of media are somewhat related to the task being accomplished, with a slight tendency to rate media more negatively when the task is more complicated and difficult. However, our results for the CSG DACOM scales indicate that the subjective ratings given by participants are much more strongly a product of the medium itself and of their degree of experience with it, and that ratings will be similar across quite different tasks. This has the effect of giving us a little more confidence in the comparative results shown in the preceding table for audio and video experiments "meaning anything", since the tasks there were different.

Chapter Seven

CONDUCTING THE EXPERIMENT: EIES AS A METHODOLOGICAL TOOL

Introduction

The computer has become an important tool for social scientists in the analysis of data from both experiments and surveys. It has also occasionally been used to conduct controlled experiments with single human subjects, in which a subject at a terminal receives pre-programmed stimuli and has his or her reactions recorded. Such use of computers in the COLLECTION of data on aspects of human behavior has been reviewed by Weiss (1973).

Recently, the computer has been introduced as a tool for the study of group or "social" behavior, rather than merely single subjects or dyads. In addition to the NJIT-based project reported here, the University of Washington has set up a "Computerized Laboratory for the Experimental Analysis of Social Interaction" (Cook and Emerson, 1977). We agree with them that the major benefit for social scientists of introducing the computer into experimentation on group processes is that it provides the "capability to expand the scope of experimental research on social interaction and to explore more fully social processes of greater complexity" (Ibid., pp. 2-3).

One major difference between the two efforts is that the computerized conferencing system as a locus for the "laboratory" makes it

available to any researcher anywhere, rather than limiting it to those who are co-located with the host computer. Another significant difference is that the experimental process can be superimposed upon an environment which can support the regular communications of a group. Thus, it can be used for extended field experiments as well as for short term laboratory experiments.

This chapter summarizes the details of how the experiment was conducted, for which full details are included in the appendices. It focuses upon the methodological aspects of using a computer program, written in INTERACT, to conduct the computerized conferencing trials. Our purpose here is to share knowledge gained about the use of the computer system as a tool for conducting such fully controlled experiments on communications processes.

The Sequence of operations

The experimental procedures and instructions were developed and refined during a pilot study and during the summer of 1978, using daytime students at Upsala as subjects. For the experiment itself, the following procedures were used:

1. Subjects were recruited by visiting classes at which a standard "recruiting speech" was presented (Appendix A).

2. Interested students were given a recruitment form to fill out. (Appendix B).

3. Potential groups were assembled by finding 7 subjects available at the same time. As many as possible were scheduled in the evenings or on Saturdays to maximize the participation of older subjects and to minimize the drain on EIES during peak hours.

4. An assistant called and made the appointment for the two sessions.

5. An assistant reminded the subjects the evening before the training or the experimental session.

6. At the training session, a standard introduction was given. (See Appendix G) After the consent forms were signed (Appendix E), assistants then took each subject to a terminal. The assistants played an essentially passive role, since all instructions were computer administered, following the guidelines for the assistant role in Appendix F.

7. The subjects spent 20 to 30 minutes receiving the instructions in Appendix W and the second instruction in Appendix X. When the assistants reported to the monitor that all subjects seemed to have mastered the commands, each subject was administered the "test' in Appendix Y. The assistants recorded performance on this test in writing. It was used as a basis for eliminating subjects whose skills were insufficient for them to be able to take part in the experiment.

8. The subjects were debriefed according to the guidelines in Appendix C.

9. On the day of the run, the subjects were thanked for returning and taken to their rooms (See Appendix G for the welcoming speech; Appendices J, K, N, and R for the sequence of monitor operations for the four conditions).

10. In the computer conferencing discussion, subjects were given a review (appendix O) before being administered the problem.

11. In face to face conditions, the subjects first read the problems in rooms by themselves. Pre-discussion Arctic rankings were obtained from each subject before they were taken to the face to face conference room.

12. After the first problem in the first mode, subjects were given a coke and cookie break. After the second problem, they completed the questionnaires comparing the two media, in their individual rooms. (See Appendix LL)

13. Subjects were debriefed (Appendix D).

The actual text of all the instructions is included in various items in the Appendix. For face to face runs, the instructions were administered orally by the Monitor or assistants. In the computerized conferencing condition, the instructions were printed out (Note: in a subsequent experiment, this has all been automated, so that once the monitor commands the experiment to start, everything is delivered to each subject at the proper time).

The discussion which follows will highlight what are seen as the methodological problems or issues in using the computer to conduct automated or partially automated experiments on group communication.

Experimenter Cueing and Prompting

With four different conditions plus the training runs with their own set of instructions for experimenter and assistants, we had a complex design to administer, subject to possible error if one had to rely on human memory. For instance, one might be conducting three runs on three nights, for which the sequence of tasks to be performed by the experimenter was totally different.

EIES was used to prompt the error-prone humans. An hour before each run, the experimenter looked in the on-line index for a list of locations of all of the instructions and operations for the particular run, then printed them out. The instruction sequences were all neatly numbered and "idiot proofed", so that even a tired experimenter would not be likely to make a mistake. The system also sent reminders to the experimenter in the computerized mode, such as "time to send final message". This final message and all other items were stored on line, so that they were exactly the same every time; the only thing that would change would be the date, time and name of experimenter shown on the top.

Data Storage On Line

All transcripts and other data produced during the computerized condition were stored permanently on line until deleted by the experimenter. This facilitates flexibility and completeness in data

collection. In addition, they are available whenever needed by members of the team, and the computer can be used to analyze some of these stored data automatically, without the time and potential errors involved in human transcription to punched cards or other media. It is possible to instrument almost any aspect of the communication being studied. For example, the transcripts were run through a routine which automatically counts and records the "number of turns" and "number of lines written" in the discussion by each participant. Routines for analyses of the various ranking data can also be run on EIES, on the stored answers that the subjects entered. For studying message traffic, a "who to whom" matrix can automatically be generated.

We have discovered advantages to having both the on-line storage of data and the complete printed original transcript of each participant. For example, we had not intended to do any coding or study of the "training" session. However, we have subsequently come up with a number of hypotheses related to the training sessions which seem worthy of testing. Since all the raw data are saved, we can go back and analyze and code for variables we did not initially intend to use.

Computer Administered Instructions to Decrease Variability

In the computerized conferencing condition, special programming was used to conduct the entire experiment with the ranking problem. This was a fairly complex series of operations, which was subject to variability and error when administered in the face to face condition. In the experiment, the problem and the initial

instructions were printed out simultaneously on all subjects' terminals. Then each subject was asked to rank-order fifteen items. The computer asked for the rank of one item at a time, then printed them back in rank order. If the subject failed to assign fifteen unique ranks to the fifteen items, the computer showed which items had been assigned the same rank, or left unranked, and asked the subject to rerank the items.

Meanwhile, as each subject completed this initial task, he or she was informed that discussion would begin whenever all five participants had completed their initial ranking. The computer automatically printed out status information to keep the subject informed of the progress of the others, such as "Two persons have now completed their ranking".

When all five had completed their ranking, a timer was set and subsequently warnings were sent to all members of the group at 45 minutes. Meanwhile, the next instruction, informing the group members that they could share their initial rankings with one another, was automatically sent. If a subject made an error in carrying out these or other instructions, the program informed the subject of the nature of the error. In other words, the computer could administer instructions that were time-triggered, triggered by an event or action by an individual, or by the completion of a specified action by a specified number of group members.

We found the computer to be much more reliable than human assistants in checking for complete data. Even though the assistants were

rehearsed and admonished to carefully check that all three sets of rankings were completed correctly by each subject, this was a complex enough task that a few mistakes were missed in the face to face condition. The presentation of stimuli to each subject by the computer is similar to many other experimental situations in pyschology. The main difference is that the computer could keep track of all five subjects simultaneously, so that each proceeded at his or her own pace.

More importantly, an experiment conducted automatically meant that unless a subject became disconnected and called for help, there was no contact with possibly biased assistants or experimenters. As Rosenthal (1966) has pointed out, the demands or attitudes of the experimenter that are implicit in non-verbal cues given to subjects can influence the results in experiments such as this one. An experiment programmed to be automatically conducted by the computer can reduce experimenter-subject direct interaction to zero, and thereby eliminate this source of bias.

Although our own experiment did involve contact with subjects during training and before and after the problem solving sessions, it is conceivable to conduct an experiment on EIES with absolutely no such contact. Russell Bernard has conducted a study in which experienced users of the EIES system who volunteered to take part NEVER interacted with the experimenter. Similarly, we have conducted one trial run of the ranking problem from our experiment using five subjects located in five different parts of the country, who likewise had no contact with the experimenter or one another except as

mediated through the computer. There is thus great promise for greater standardization than is possible with other techniques for experimenting with groups.

Given that EIES now has a population of over five hundred users engaged in professional communications, the use of short term subjects for specific experiments can be supplemented with regular users of the same communication medium. This provides a supplemental form of control in establishing boundaries on the range of generalizability of the results to other types of subjects.

Problems

Although we think that the use of computerized conferencing is a promising new tool for experimentation in group processes, we would not like to leave the impression that there are no difficulties in its use at this stage of its development. The most serious of these is the training of subjects. Even though we were using college students who had claimed that they could at least "hunt and peck" on a typewriter, it turned out that some of them had to hunt a minute or so in order to find a single key to peck; or that they did not have good enough command of written English to be able to communicate effectively in writing. They had to be eliminated from the experiment if we wished to have five actual participants in each problem solving discussion.

In addition, the training took about an hour, and was so draining that the subjects were incapable of spending two hours on a difficult problem afterwards. Fear and doubt about their ability to use such a

"complicated" thing as a computer terminal seemed to raise their mental effort. After about twenty to thirty minutes, almost all trainees were declaring that they were at least "comfortable" with the terminal and understood how to use the system to communicate with one another; many were claiming by the end of half an hour that it was even "fun". However, by this time, the letdown from the state of high anxiety had taken its toll. Therefore, the subjects should be trained in one session before they can be used in an experiment in a subsequent session. Ideally, subjects would participate in several problems or experiments after training, in order to maximize the return on this investment for the experimenter. An alternative approach is the one which we took in our subsequent experiment-- the participants were given a full lunch break after training, followed by only a single problem.

Chapter Eight

Summary and Conclusion

We found the following differences in comparing face to face (FtF) with computerized conferences ("CC") for two kinds of tasks that had associated differences in the structuring of the interaction:

1. There was no difference in the quality of solution reached for the ranking problem, which had an expert criterion solution.

2. Both FtF and CC groups improved about 25%. The majority of groups produced better solutions than those held by any of the members before the group discussion.

3. For the qualitative human relations task, there was some tendency in CC for decisions to be more reward-oriented and less punitive.

4. For the unstructured, value laden problem, there was a very striking difference in the ability of the groups to reach total consensus. All eight of the FtF groups reached consensus on this problem, but only one of the CC groups did.

5. On the structured ranking problem, which was a knowledge-sharing task, FtF groups were also more able to reach total consensus (half did), but all of the CC groups reached at least 80% consensus.

6. There were two to three times as many units of communication in

face to face conferences as in the computerized conferences.

7. There were many differences in communications process as profiled by Bales Interaction Process Analysis. These differences were a function of task type as well as of medium.

8. Many of the differences in interaction process are significantly related to the ability of a group to reach consensus and/or reach a high quality decision. These relationships between communications processes as measured by Bales Interaction Process Analysis and communications outcome are somewhat different for the two communications modes.

9. In the more structured, knowledge-sharing task, there were no differences in inequality of participation or dominance between media. For the unstructured , value-laden task, there was notably more tendency for a single dominant person to emerge in the discussion in the FtF condition.

10. Though CC as a mode of communication received generally satisfactory ratings, face to face communication was felt to be significantly more satisfactory by the participants in this experiment. Comparisons with the ratings on the same scales by long term users of CC suggest that subjective satisfaction may be largely a function of amount of experience with the medium.

Generalizability of Results

The conditions under which the participants in this experiment used computerized conferencing were far from optimum. A higher baud rate than the 30 characters per second used would be more satisfying-- of course, this requires a special modem and high quality telephone lines, and/or a more expensive terminal than the "dumb" printing terminal we used in this experiment. Any organization which installed its own conferencing system would undoubtedly invest in the equipment necessary to provide a higher speed than we used.

The computerized conferencing mode of communication used in this study was perhaps the most adverse set of circumstances. Inexperienced participants who had never met or worked with one another previous to the experiments were under considerable time pressure in an unfamiliar medium to come up with a solution. Moreover, the medium was being used synchronously, whereas its strengths are more apparent in an asynchronous condition, when each individual participates at a time of their own choosing and can get off line and think and look up references to help them formulate their contributions. It is probable that experienced users in "real" groups employing an asynchronous pattern of use would be more effective and more satisfied with the medium.

In addition, there was no attempt to use the power of the computer to provide feedback to the group or to provide a structure for their discussion. We think that this can be more effective than "free

discussion" formats for computerized conferencing.

Therefore, we feel that the results we obtained are very conservative-- in the sense that there are many changes that could be made in the details of the way in which the computerized conference was conducted that would probably improve the process and outcome in relation to face to face discussions. We feel that we have demonstrated, however, that even very inexperienced users of a very simple, low speed system, can participate in a group discussion, and that the outcome is likely to be as good in terms of quality of decision reached as if they had met face to face. The various enhancements and improvements that could be made to the simple form of CC should enable it to perform "better than" FtF conferences on some dimensions, while it will remain "worse" on others.

Design of the Next Experiment

Experiment Two has been designed to explore the possibility of using human leadership or computer decision aids to improve the quality of decisions and the ability to reach consensus in a computerized conference. It is a two-by-two factorial design, with the factors 1) presence or absence of an elected human leader with a defined leadership role; and 2) presence or absence of a computer feedback table which analyzes the individual decisions and shows the areas of agreement and disagreement.

In order to provide indirect comparison to the results of Experiment One, the rank ordering problem called "Lost in the Arctic" is being used again. However, several changes were made, even in the condition which essentially replicated the unstructured conference in experiment one. The changes are:

1. All groups are actual groups of five individuals from organizations, and not students coming to a laboratory. Thus, this is a field experiment, brought to the offices of the participating persons. The participants will have a common organizational identity.

2. Since the computer conference groups in Experiment One seemed rushed by a 90 minute time limit in which to make a decision, the groups in Experiment Two are being given a two hour time limit.

3. The Experiment Two groups were given a longer practice period, including a practice ranking problem.

4. A command was created (+order) which enabled the participants to have their current rank order displayed, and to change this order at any time; other group members were automatically notified of such changes as they were made.

5. Approximately every ten minutes, a table displaying the raw data showing the current rank orders of the five members was created and displayed to all members.

6. The experiment, including the training, was completely automated, and a much fuller record of the details of each participant's actions was logged automatically.

7. The interface was somewhat simplified, with only four commands and no menu choices.

The third experiment in the series will be a set of long-term field experiments in organizational settings. Thus, as the series of experiments progresses, we are moving further away from the highly controlled but oversimplified conditions of experiment one, and closer to studies of variations in the computerized conferencing mode within "real world" organizational settings. In doing so, we will sacrifice the extent to which we can determine "cause and effect" among a complex set of variables, but will be able to determine the extent to which laboratory findings seem to be generalizable to operational settings.

REFERENCES

Bales, Robert
1950 Interaction Process Analysis: A Method for the Study of Small Groups. Reading, Mass.: Addison Wesley.

Bales, Robert F.
1955 "How People Interact in Conferences". Scientific American, 192, 3, 31-35.

Bales, Robert F. and Edgar F. Borgatta
1955 "Size of Group as a Factor in the Interaction Profile". In A.P. Hare, E.F. Borgatta and R.F. Bales, eds., Small Groups: Studies in Social Interaction, pp. 396-413. New York: Knopf.

Bales, Robert, F.L. Strodtbeck, T.M. Mills and M.E. Roseborough
1951 "Channels of Communication in Small Groups". American Sociological Review, 16, 461-468.

Bales, Robert F. and F.L. Strodtbeck
1951 "Phases in Group Problem Solving". Journal of Abnormal Psychology 46, 485-495.

Berger, J., B.P. Cohen and M. Zeldich, Jr.
1966 "Status Characteristics and Expectation States". In J. Berger, M. Zeldich and B. Anderson, eds., Sociological Theories in Progress, Vol. 1, 29-46. Boston: Houghton Mifflin.

Berger, J., B.P. Cohen and M. Zeldich, Jr.
1972 "Status Characteristics and Social Interaction". American Sociological Review 37, 241-255.

Berger, J. and M.H. Fisek
1974 "A Generalization of the Theory of Status Characteristics and Expectations States". In J. Berger, M.J. Fisek and T.L. Connor, Expectation States Theory: A Theoretical Research Program, 163-205. Cambridge: Winthrop Publishers, Inc.

Berger, J., M.H. Fisek, R.Z. Norman and M. Zeldich, Jr.
1976 "Status Characteristics and Social Interaction: An Expectation States Approach". New York: Elsevier-Noerth Holland.

Burke, Peter
1974 "Participation and Leadership in Small Groups". American Sociological Review 39, 832-843.

Carlston, Donald E.
1977 "Effects of Polling Order on Social Influence in

Decision-Making Groups". Sociometry 40, 2, 115-123.

Chapanis, A. and C. Overby
1974 "Studies in Interactive Communication, III". "Effects of Similar and Dissimilar Communication Channels and Two Interchange Options on Team Problem Solving". Perceptual and Motor Skills 38, Monograph 2-v38.

Chapanis, Alphonse
1975 "Interactive Human Communication". Scientific American, 232,3, 36-42.

Chappel, E.A, and M. Arensburg
1948 "Measuring Human Relations: An Introduction to the Study of the Interaction of Individuals". Genetic Psychology Monographs 32, 3+147.

Cohen, E.G. and S.S. Roper
1971 "Modification of Interracial Interaction Disability: An Application of Status Characteristics Theory". American Sociological Review 37, 643-57.

Cook, Karen S. and Richard M. Emerson.
1977 "Design of a Computerized Laboratory for the Experimental Analysis of Social Interaction". Technical Report, Department of Sociology, Univ. of Washington.

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

Eskilson, Arlene and Mary Glenn Wiley
1976 "Sex Composition and Leadership in Small Groups". Sociometry 3, 183-194.

French, R.L.
1950 "Verbal Output and Leadership Status in Initially Leaderless Group Discussions". American Psychologist 5, 310.

Hackman, J. Richard and Charles G. Morris
1975 "Group Tasks, Group Interaction Process, and Group Performance Effectiveness: A Review and Proposed Integration". Advances in Experimental Social Psychology, Vol. 8. Reprinted in Berkowitz, Leonard, Ed., Group Processes. New York, San Francisco and London, Academic Press, 1978, pp. 1-55.

Hiltz, Starr Roxanne
1975 "Communications and Group Decision Making": Experimental Evidence on the Potential Impact of Computer Conferencing. Newark, N.J., Computerized Conferencing and Communications Center, New Jersey Institute of Technology, Research Report No. 2.

Hiltz, Starr Roxanne
1976 "A Social Scientist Looks at Computerized Conferencing", Proceedings, Third ICCC, Toronto.

Hiltz, Starr Roxanne, Kenneth Johnson and Gail Agle
1978 "Replicating Bales Problem Solving Experiments on a Computerized Conference: A Pilot Study". Computerized Conferencing and Communications Center, New Jersey Institute of Technology, Newark, N.J., Research Report No. 2.

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

Johansen, Robert, R. DeGrasse, and T. Wilson
1977 Group Communications Through Computers, Vol. 5, Effects on Working Patterns. Institute for the Future, Menlo Park, Cal.

Johansen, Robert, Jacques Vallee and Kathleen Spangler
1979 Electronic Meetings: Technical Alternatives and Social Choices.

Kanter, Rosabeth Moss
1977 "Some Effects of Proportions on Group Life": Skewed Sex Ratios and Responses of Token Women". American Journal of Sociology 82,5: 965-990.

Katzell, R.A., C.E. Miller, N.G. Rotter and T.G. Venet
1970 "Effects of Leadership and Other Inputs on Group Process and Outputs". Journal of Social Psychology, 80, 157-169.

Meeker, B.F. and P.A. Weitzel-O'Neill
1977 "Sex Roles and Interpersonal Behavior in Task-Oriented Groups". American Sociological Review 42,1: 91-104.

Merton, Robert
1968 Social Theory and Social Structure. (Revised and enlarged edition). New York: The Free Press.

Milton, G.A.
1959 "Sex Differences in Problem Solving as a Function of Role Appropriateness of the Problem Content". Psychological Reports 5: 705-8.

Norfleet, B.
1948 "Interpersonal Relations and Group Productivity". Journal Of Social I3sues 2, 66-69.

Osborn, A.
1963 Applied Imagination, Third Edition. New York:

Scribner.

Parsons, Talcott, R.F. Bales, and E.A. Shils
1953 Working Papers in the Theory of Action. Glencoe: The Free Press.

Pye, Roger and Ederyn Williams
1977 "Teleconferencing: Is Video Valuable or is Audio Adequate?". Telecommunications Policy, June, 230-241.

Rosenthal, R.
1966 Experimenter Effects in Behavioral Research. New York: Appleton-Century-Crofts.

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

Slater, P.E.
1955 "Role Differentiation in Small Groups". American Sociological Review 20: 300-310.

Shaw, Marvin E.
1976 Group Dynamics: The Psychology of Small Group Behavior. McGraw-Hill, second edition.

Shaw, M.E.
1963 "Scaling Group Tasks: A Method of Dimensional Analysis". JSAS Catalogue of Selected Documents in Psychology 3,8 (MS No. 224).

Strodtbeck, F.L., R.M. James, and C. Hawkins
1958 "Social Status in Jury Deliberations". In E.F. Maccoby, T.M. Newcomb and E.L. Hartley, eds., Readings in Social Psychology 3rd edition. New York: Henry Holt and Co.

Strodtbeck, F.L. and R.D. Mann
1956 "Sex Role Differentiation in Small Groups". American Sociological Review 20: 300-310.

Torrence, E.P.
1954 "Some Consequences of Power Differences on Decision Making in Permanent and Temporary Three Man Groups". pp. 130-140 in Research Studies 22. Pullman, Washington: Washington State University.

Vallee, Jacques, Robert Johansen, Hubert Lupinski, Kathleen Spangler and Thaddeus Wilson
1975 Group Communication Thrkugh Computers, Volume 3: Pragmatics and Dynamics. Menlo Park, Cal., Institute for the Future, Report R-35. Reading, Mass., Addison Wesley Publishing Co.

Vallee, Jacques, Robert Johansen, Robert Randolph and Arthur Hastings
Group Communication Through Computers: Volume 2, A Study of Social Effects. Menlo Park, Cal., Institute for the Future, R-33.

Van De Ven, Andrew and Andre L. Delbecq
1974 "The Effectiveness of Nominal, Delphi and Interacting Group Decision Making Processes". Academy of Management Journal 17, 4, 605-621.

Vroom, Victor L., Lester D. Grant and Timothy S. Cotton
1969 "The Consequences of Social Interaction in Group Problem Solving". Organizational Behavior and Human Performance 4, 77-95.

Webster, Murray, Jr.
1974 "Actions and Actors". Cambridge: Winthrop Publishers Inc.
1977 "Equating Characteristics and Social Interaction: Two Experiments". Sociometry 40, 1: 41-50.

Weiss, Bernard
1973 Digital Computers in the Behavioral Laboratory. New York: Appleton-Century-Crofts.

Werner, C. and B. Latane
1976 "Responsiveness and Communication Medium in Dyadic Interactions". Bulletin of the Psychonomic Society 8, 11: 13-15.

Westgate, T.N.
1978 Group Communication and Conferencing Through Computers: The Implications for the Management of Organizations. Research Reporp No 1, Mount Allison University, New Brunswick, Canada.

Willard, Don and Fred L. Strodtbeck
1972 "Latency of Verbal Response and Participation in Small Groups". Sociometry 35,1: 161-175.

Williams, Ederyn
1975 "Coalition Formation Over Telecommunication Media". European Journal of Social Psychology 5,4: 503-507.