AN EMPIRICAL COMPARISON OF COPRESENT AND TECHNOLOGICALLY-MEDIATED INTERACTION BASED ON COMMUNICATIVE BREAKDOWN

by

ECKEHARD DOERRY

A DISSERTATION

Presented to the Department of Computer and Information Science
and the Graduate School of the University of Oregon
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

December 1995

ABSTRACT

Within the area of Computer-Supported Cooperative Work (CSCW), there has been an explosion of interest in how recently developed network technologies might be applied to support the collaborative endeavors of widely distributed participants. Increasingly powerful systems for desktop conferencing, group meeting, and distributed design have been developed. Though the technologies applied in such systems vary widely, their underlying design goal is essentially the same: to support interactions that are functionally equivalent to face-to-face interaction.

This dissertation evaluates the extent to which currently available technologies achieve this goal by comparing the amount of *communicative breakdown* experienced by pairs of participants interacting in three communication environments: copresent, audio-mediated and audio/video-mediated. In all three environments, participants had access to a shared workspace, in which they used a graphical computer simulation to collaboratively explore the behavior of a simple cardiovascular system.

Videotaped interactions were analyzed in a series of three studies, intertwining the qualitative techniques of Conversation and Interaction Analysis with more traditional quantitative techniques to progressively refine understanding of the functional differences that exist between environments. Four categories of communicative breakdown were identified: failure to maintain shared conceptions of current topic, failure to establish shared reference, and failure to regulate access to the verbal channel and to a shared cursor.

Statistical results showed that copresent interactions were significantly less prone to breakdown than interactions in either of the two technologically-mediated environments; no significant differences in the incidence of breakdown were found

between audio-only and audio-video interactions. A subsequent qualitative analysis showed that breakdowns in technologically-mediated interactions were related to a profound insensitivity to nonverbal displays like direction of gaze, deictic gesture and manipulation of objects in the task context. This result demonstrates that, though visual access to a partner is clearly vital for avoiding breakdown, the visual access afforded by a video image is fundamentally unequal to that afforded by physical copresence. More generally, there is a great deal of difference between *technically* making more communicative resources available in an environment and the *practical* utility of such upgrades to participants.

ACKNOWLEDGEMENTS

The author expresses sincere appreciation to Professor Sarah Douglas for her assistance in the preparation of this manuscript. Without her advice and steadfast support of this project, this work could never have been accomplished. Thanks are due to Professor Jack Whalen as well, whose special familiarity with the concepts and techniques of ethnomethodological research was crucial in shaping the research presented here. Discussions with Marilyn Carter and Kartik Mithal yielded valuable insights over the course of this project. Thanks also to Chris Hundhausen and Ted Kirkpatrick for their meticulous editing help.

DEDICATION

This dissertation is dedicated to my parents, Karl and Brigitte Doerry, without whose continuous support, encouragement, and enthusiasm for higher learning the work presented in this dissertation would never have been undertaken, and to my wife, Sharon, whose quiet strength and endless patience sustained me during those times when nothing appeared to make sense.

TABLE OF CONTENTS

Page

CHAPTER I

TECHNOLOGICALLY-MEDIATED INTERACTION

From the very beginnings of the computer age, the potential of computers as tools for supporting everyday human activities has figured prominently in the collective vision of both the public and the research community. The capability to flawlessly store, recall and process massive amounts of data with lightning speed distinguished the computer processor from the fallible human mind, and appeared to make it a perfect mechanism for managing and transmitting the rapidly increasing volume of human knowledge to new generations. At the same time, it is unclear what roles computers can and should play in supporting human endeavors. A central theme of research in the four decades that computer science has existed as a distinct discipline has been the exploration of this issue. What tasks are computers capable of performing and what is their relationship with human users in that task-solution context?

Until very recently, many research efforts within computer science were driven by the conviction that "intelligent machines" could interact with humans as equal partners, performing human-like reasoning tasks, discussing problems and solution strategies, and even taking over certain knowledge-based activities entirely. For example, if computers could be made to play the role of teachers, the entire educational process could be revolutionized, with students receiving individualized instruction tailored to their unique learning styles. More generally, the vision was one in which computers imbued with the knowledge of human "domain experts" would perform the activities of those experts flawlessly and efficiently, vastly expanding access to expertise and freeing humans to devote their energies elsewhere. Computer-based replacements for human expertise have been explored in a broad variety of domains including classroom instruction, internal medicine, VLSI design, geochemical engineering and many others.

In the last decade we have seen the boundless enthusiasm for intelligent machines founder, as system after system has failed to live up to expectations. Of all the systems and approaches that have been suggested, very few have ever found their way into the real world, and almost none of these have been used as their designers originally intended, namely, as stand-alone replacements for human expertise.

In response to these difficulties, a new and vibrant area of research has recently emerged within the computer science community. The area of Computer-Supported Cooperative Work (CSCW) is dedicated to exploring the ways in which machines can support the collaborative interactions of human users. This change in research focus marks a fundamental reassessment of the role that computers are expected to play in human society, backing away from the ambitious goals of artificial intelligence, and shifting the computer into a supportive rather than participatory role in human problem-solving activities. In this way, CSCW represents a gradual shift of interest within the research community away from the vision of the intelligent machine, with its attendant focus on the computerized representation, manipulation, and transfer of abstract symbolic knowledge, and towards a more social conception of computing, focused on the way in which computers might support collaborative interactions between humans.

In an ideal world, all collaborative interactions would take place between copresent participants; the copresent condition clearly represents the most natural collaborative context, allowing participants to draw directly on a lifetime of communicative experience to organize their interaction. Unfortunately, the material and geographic constraints of the modern world make personalized interactions of this sort increasingly unlikely. Work groups may be distributed across widely separated subsidiaries of a large organization and may run into the tens or hundreds of participants. Accordingly, an area within CSCW that has received much attention in recent years is exploring the ways in which computer-based technologies can support the collaborative interactions of users that are geographically distributed. In general, the goal is to create powerful electronic communication environments that can serve as substitutes for copresent interaction, allowing users to accomplish their communicative and creative goals without having to be physically in the same place. By greatly reducing the expense and physical effort of communicating with a collaborator, electronic communication environments herald profound changes in the ways we work and communicate with each other, fundamentally reshaping the dynamics of social interaction in modern society. Communities of practice would no longer be constrained by geographical proximity, allowing members to meet and work collaboratively in the virtual medium defined by the electronic environment. For instance, business partners could meet in sophisticated audio-video environments that allow them to see and interact with each other as they collaboratively edit business documents represented a shared electronic workspace; university classes could be held in electronic classrooms that bring together students and teachers from around the globe, allowing participants to interact and accomplish collaborative work, both personally or as larger groups; members of research communities could meet in virtual forums that allow widely-distributed participants to present data in a mutually available electronic space, and support naturalistic discussion and manipulation of such data.

These utopian visions of substituting technologically-mediated interaction for copresent interaction have resulted in the development of an exceedingly broad variety of electronic communication environments, ranging from simple systems for organizing text-based email interaction to costly virtual reality environments that create entirely artificial communicative contexts. Though the technologies that are applied in these systems vary widely, the underlying goal of all such systems is essentially the same: to provide a simulacrum of copresence that is somehow functionally equivalent to copresent interaction, allowing participants interacting in technologically-mediated environments to communicate and collaborate just as effectively as if they were physically copresent. In other words, the goal of any electronic environment is to provide the same *communicative efficacy* as face to face interaction.

The basic issue explored in this dissertation is the extent to which existing technologically-mediated communication environments ever truly achieve this goal. Specifically, this research is aimed at addressing the following issue:

> How good are the simulations of copresence embodied in existing systems? That is, how does the communicative efficacy of these distributed environments compare to true copresent interaction?

Clearly, the only way to answer this question is to somehow evaluate and compare the communicative efficacy of copresent and technologically-mediated interactions. This raises a second, more pragmatic question:

> How can the notion of communicative efficacy be operationalized? What features of an environment or interactions occurring within that environment should serve as metrics for assessing the communicative efficacy of the environment?

The underlying issue raised by these questions is one of evaluation. The final step in the design of any engineered artifact, from a simple mousetrap to the most advanced space shuttle, is to somehow evaluate the performance of the artifact with respect to the original design goals. Only by articulating the extent to which the designed artifact satisfies these goals can the success or failure of a design be meaningfully established.

The central motivation for the research presented in this dissertation is that we currently have no basis for understanding the extent to which technologically-mediated interaction is functionally equivalent to copresent interaction, because the way in which the performance of existing technologically-mediated communication environments is evaluated is fundamentally flawed. The analysis provided in this work remedies this shortcoming by empirically comparing the communicative efficacy of technologically-mediated and copresent interaction. Drawing on theoretical and methodological foundations recently developed in the social sciences, a methodology for comparing the communicative efficacy of environments based on the number of "communication breakdowns" experienced by interacting participants is developed, and then applied to compare the efficacy of copresent interaction to that of interactions in two representative distributed environments.

A key advantage to using communicative breakdown as a metric for assessing communicative efficacy is that it yields a concise articulation of what is going wrong in the communicative interaction of collaborators, providing a strong basis for explaining *why* observed differences in communicative efficacy exist. In this way, the analysis is able to inform the design of future systems. More importantly, this analytic approach exposes fundamental limitations associated with the technologies used to simulate copresence in electronic environments that suggest that environments relying on these technologies are unlikely to ever provide the same communicative efficacy as copresent interaction.

To establish a foundation for the research presented in this dissertation, an extensive survey of technologically-mediated communication environments is presented in the following section. Later sections critically examine current approaches to evaluating and comparing the communicative efficacy of these environments, revealing profound deficiencies in existing evaluative techniques and motivating the development of a more powerful analytic tool to serve as the methodological cornerstone of the comparative analysis undertaken in this research.

1.1 CSCW: Technologically-Mediated Collaboration

The goal of Computer-Supported Cooperative Work (CSCW), as a research area, is to explore ways in which computers can be used to manage the vast complexity of interaction in large and/or distributed groups. Research interests in CSCW are broadly distributed, ranging from the analysis of group dynamics and processes to technically oriented efforts to develop networks, voice applications, co-authoring tools, shared databases, and decision support systems.

The following sections survey recent developments in the field of CSCW, briefly describing a variety of systems that have been developed in the last five years to support the collaborative activities of distributed participants. The survey is organized by drawing a rough distinction between systems based on the overall nature of the interactions they are designed to support. Systems designed for *personal interaction* are primarily aimed at supporting the mundane social contacts that constitute the bulk of our everyday interactions with others. Examples of such interactions include checking to see if someone is in his or her office, contacting a friend for lunch, or having a conversation with a business associate. By contrast, *task-oriented* systems are designed to support specific collaborative problem-solving activities that yield a tangible result or solution. The primary goal of these systems is to support electronic representations of the problem statement or its evolving solution that are as robust as the representational mechanisms available to copresent participants. Though many systems in this category have focused on shared sketching and drawing, other examples include collaborative browsing of a database, group authoring, and collaborative data analysis. Table 1.1 summarizes the distinction between these two classes of systems and gives an overview of existing systems that fall within each class.

Table 1.1: Comparison and overview of systems designed to support personal versus task-oriented interactions.

	Communicative activities emphasized by system	Examples of Existing Systems
Systems for Personal Interaction	• Desktop conferencing • Personal messaging • Casual social contact • Group meetings	• COORDINATOR • CRUISER • LambdaMoo • CAVECAT • Portholes
Task-Oriented Systems	• Shared drawing/sketching • Group authoring • Decision support • Collaborative Design	• GROVE • Commune • Mediaspace • VideoWhiteboard • Clearboard

It is important to emphasize that the distinction between these two types of systems merely provides a rhetorical framework for structuring the upcoming survey. In particular, there is no implication that systems designed for personal interaction are never used to mediate a task-based activity, or that task-oriented systems are never used for non-task-related interactions. In practice, most systems provide at least some support for both facets of collaborative interaction; they are categorized based on which of these two activities the design effort primarily aims to support.

1.1.1 Supporting Personal Interaction

A wide variety of technologies have been explored in an effort to support the mundane personal interactions of widely-distributed conversational partners as they engage in the everyday activities of institutional life. The following sections survey a representative sample of systems based, respectively, on interactive text, interactive audio, interactive video, and virtual reality.

1.1.1.1 Typed Text: Maximally Constrained Interaction

Text representation is compact and already exists as a primary symbolic representation supported by computers, making the sending of text between machines a natural and simple extension. An interesting issue is raised by the asynchronous nature of typed-text interaction: Can typed-text interactions be considered interactive communication despite the fact that they are chronologically disjoint? In other words, where is the dividing line between individually constructed narratives like books or letters and collaboratively constructed "interactive" conversation. Some systems (Comer & Peterson, 1986; Kaplan, 1990; Shepherd, Mayer, & Kuchinsky, 1990) clearly consider message-based interactions like email exchanges to be interactive conversation. Taking this idea to an extreme, COORDINATOR™ (Action Technologies, 1989) explicitly requires participants to classify their messages into categories (e.g., request, commitment, etc.) defined by Speech Act theory (Austin, 1962; Searle, 1979) .

More recently, increasing connectivity and reliability of wide area networks has led to an explosion in interactive text communications. Simple systems like the UNIX talk program merely provide a "textual telephone," allowing one user to contact another, after which they correspond by typing characters into a mutually available text space. The Internet Relay Chat (IRC) program (Oikarinen, 1988) provides a large number of such textual forums, allowing users to select the one they wish to participate in. Though such applications have found some success in contexts where telephone contact is impractical or expensive (Reid, 1992) , the overall acceptance of such systems has been lukewarm. Grudin's (1988) suggestion that acceptance of a CSCW tools turns on the effort/benefit ratio perceived by users is clearly applicable here: merely providing a textual substitute for telephone conversation requires considerable additional effort with limited benefits.

Multi-User Dungeons (MUDs) are an example of typed-text environments that provide substantial extensions to the telephonic metaphor which, judging by the recent explosion in MUD popularity, more than compensate for any added typing effort. Rather than simply allowing two-participant conversations, MUDs allow multiple users to engage in (textual) interaction. As an organizational framework, the metaphor of multiple rooms (i.e. logically distinct electronic spaces) was adapted from text-based adventure games (c.f. Bartle, 1990; Evard, 1993) . Briefly, the MUD system defines the structure[1] of the virtual space and distributes the input of connected users appropriately. Each user is embodied within the virtual space as a "character," and can cause that character to speak and act in various ways.

The fact that they define a communication environment that exists independently of the conversations that go on within that environment distinguishes MUDs from most other simulations of copresence. The MUD defines not only a communication channel, but also a (virtual) "place" to meet. Indeed, some of the most interesting aspects of MUDs center around the social phenomena that develop in such forums (Curtis, 1992; Kiesler, Siegel, & McGuire, 1991) .

1.1.1.2 Audio Links: Well-Understood Technology

Audio linkages are perhaps the oldest means of technologically-mediated interaction, beginning with the telegraph and evolving into today's modern telephone networks. The fact that little effort has been invested in exploring computer-supported audio connections does not mean that audio is considered unimportant. On the contrary, several studies (Chapanis, 1975; Oviatt & Cohen, 1989) emphasize that audio contact is the single most important resource for collaborative interaction. A more likely explanation for this lack of interest is that the pervasiveness and robustness of modern telephone technology makes such work largely redundant. Nonetheless, at least one project has explored the computer's potential for overcoming traditional limitations of telephonic communication. The recently developed VAT protocol (Lawrence Livermore Labs, 1992) defines over 300 network-based channels, each of which serves as a forum for an unlimited number of participants. When users connect to a given channel, their audio input is continuously combined with the audio input of all other channel subscribers, with the resulting aggregate distributed to all subscribers.

In practice, audio channels are more commonly used to support task-oriented aspects of collaboration. For instance, in the GROVE multi-user authoring system (Ellis, Gibbs, & Rein, 1991) , users communicate by audio while simultaneously editing a document in a shared workspace. Similarly, other projects (Bly & Minneman, 1990; Minneman & Bly, 1991; Tang & Minneman, 1990; Tang & Minneman, 1991) connect users with an audio link while they collaboratively modify a shared drawing space.

1.1.1.3 Multimedia Environments: Combining Audio and Video

Looking at the literature, it is clear that interest in providing a visual connection between distributed participants has increased dramatically in recent years. The motivation for such systems is two-fold: First, there has been a general assumption[2] that providing a video channel inherently improves the communicative efficacy of a technologically-mediated environment. Second, there are indications (Kraut, Fish, Root, & Chalfonte, 1993) that frequent informal contact is a key to productive group work. That is, conversations resulting from chance encounters at the drinking fountain may be just as important to group productivity as planned group meetings. This second observation in particular is responsible for a host of systems based on the notion of browsing through a virtual space populated by audio/video connections to other offices or public places within the organization. For instance, CRUISER (Root, 1988) places video camera and microphone in each group member's office as well as in hallways and meeting rooms. Participants are able to define "pathways" through this space, moving (hence the system's name) continually between the connections placed on the path. As an example, one could check to see if a colleague is in his or her office and seems to be free, and then start a conversation with that person. CRUISER arguably defines the pinnacle of technical refinement, providing live-frame video and high fidelity audio connections, along with innovative tools for establishing and managing connections. Projects with similar goals and arrangements abound: CAVECAT (Mantei, Baecker, Sellen, Buxton, & Milligan, 1991) provides audio and video channels between as many as four sites; Portholes (Dourish & Bly, 1993) and Polyscope (Borning & Travers, 1991) provide a snapshot images of selected workspaces, updating them at predefined intervals; MMCC (in conjunction with VAT and NV) (Lawrence Livermore Labs, 1992) provides tools to organize network multicast of audio and video. The VROOMS (Borning & Travers, 1991) system slightly modifies the above formula, by defining the notion of virtual rooms, electronic spaces where people can meet.

Another class of systems is centered around the concept of casual interaction between large distributed groups. VideoWindow (Fish, Kraut, & Chalfonte, 1990) works to provide the illusion that two widely separated group meeting rooms are actually adjacent. The copresent group in each room has a large window (i.e. a screen) into the other room, showing the action in that room. Audio connections exist, allowing distributed groups to meet at the window and carry on a conversation. The SCL project (Abel, 1993) combines the group meeting with the personal meeting concept, supporting both a group meeting space and connections between individual offices.

Though most multimedia environments can be considered to by task-oriented in some sense, certain systems provide express support for accomplishing collaborative work that goes beyond social communication. Specifically, task-oriented[3] systems provide a dedicated channel for electronically representing the evolving solution to a task that participants are working on. Mediaspace (Harrison & Minneman, 1990) , for instance, is designed specifically to support collaborative design activity between distributed participants. Multiple cameras and monitors were placed in each participant's work area, with computers coordinating the connections between work areas. The Teamworkstation (Ishii, 1990) project is based on the notion that easy access to the evolving problem-representation is just as important as visual access to other participants in the interaction. Accordingly, the shared drawing area and visual images of remote participants appear in the same workspace. A similar approach is taken in (Dykstra-Erickson, Rudman, Hertz, Mithal, Schmidt, & Marshall, 1995) .

1.1.1.4 Virtual Reality

Unlike the simulations of copresence discussed above, which provide some sort of audio or visual "window" into the remote contexts of other conversational participants, Virtual Reality (VR) systems establish a completely artificial communicative context created and maintained by the system. The distinction being drawn here has profound implications for the notion of simulating copresence. The sense of copresence in text, audio, and video linkages is based on the *distribution* (i.e. transmission) of every participant's context to all other participants; the shared communication environment is forged from the patchworked union of all individual contexts. In contrast, VR systems are based on the notion of virtual *displacement*, removing users from their individual physical contexts and bringing them together in a shared virtual communicative context.

As a nascent technology, much work in VR remains focused on developing appropriate control structures for virtual interaction by, for example, extending the concept of the UIMS developed for graphical event-driven desktop interfaces to manage input from devices like the VPL Dataglove™ and the Polhemus 3Space™ tracker (Lewis, Koved, & Ling, 1991) . Moving up a level, others (Mackinlay, Robertson, & Card, 1991) have worked to develop control metaphors for navigating through large three-dimensional spaces.

In sum, VR is a technology with much apparent potential. While the costs of such systems are, relatively speaking, still in the stratosphere, they are dropping rapidly (Pausch, 1991) . The real question, however, is not whether one can afford VR, but whether it constitutes a more robust sense of copresence than, say, an audio/video link. As discussed above, VR takes the notion of copresence to a higher level, by working to remove participants from their real world contexts and place them in a shared virtual context. While this obviates the problem of accurately transmitting participants' real contexts, it places the onus of creating and maintaining a complete virtual context entirely on the system. Given evidence that humans rely on detailed and multi-faceted aspects of context to organize their communicative behavior, this may be a leap from the frying pan into the fire. Indeed, most VR systems do not attempt to model copresence by allowing interaction between multiple participants within the virtual space. Existing applications have mainly explored the utility of VR as a way of viewing large data spaces (Card, Robertson, & Mackinlay, 1991; Robertson, Mackinlay, & Card, 1991) and complex simulations (Lewis, Koved et al., 1991) .

1.1.2 Supporting Task-Oriented Interactions

The emphasis in the design of systems for supporting task-oriented interactions is on the design of electronic workspaces that allow participants to represent and manipulate their evolving solutions over the course of their interaction. The most primitive instances of such systems simply allow multiple users to simultaneously access and modify a document from their individual workstations, or to regulate shared access to some dataspace within the system. At the other extreme are systems that work to extend the functionality of the advanced audio/video environments discussed in Section 1.1.1, by providing an electronic space in which participants share access to some representation of the problem they are working on.

In the CSCW community, the utility of sharing electronic data spaces between widely separated users has long been recognized (Engelbart, 1975) . Interest in the last decade has grown substantially, resulting in both experimental and commercial development efforts. Lauwers and Lantz (1990) divide such efforts into two categories, based on the amount of problem-specific support provided by the system for collaboration. Collaboration-aware systems (Ellis, Gibbs et al., 1991; Stefik, Foster, Bobrow, Kahn, Lanning, & Suchman, 1988) are designed around applications specifically designed for use by multiple users. Clearly, this approach is advantageous in that it allows designers to incorporate application-specific tools for distributed collaboration. A more economical approach (Ahuja, Ensor, & Lucco, 1988; Gust, 1989) allows existing single-user applications to be shared by multiple participants by providing multiplexing mechanisms at the UIMS level. In such systems, the operating system allows any display area (i.e. window) to be selectively shared by multiple participants. Individual participants may take part in multiple simultaneous collaborations, sharing certain windows with each collaborative group.

Within the overall effort to share data spaces, two classes of collaborative activity have drawn an unusual amount of attention: shared drawing and group authoring. Shared sketching and drawing has proven difficult to support for several reasons (Minneman & Bly, 1991) . Potentially large graphical spaces and the marks and gestures made in them must be distributed to all participants with minimal delay. At the same time, marking, gesturing and erasing must be simultaneously enabled for all participants. Bly and Minneman's Commune (1990) system meets these challenges, but provides only a limited sense of gesture: Only those gestures that occur *on* the drawing tablet are conveyed to the remote site. LiveBoard (Weiser, 1991) provides similar functionality, but uses a vertical whiteboard rather than a drawing tablet. Videodraw (Tang & Minneman, 1990) and its successor, VideoWhiteboard[4] (Tang & Minneman, 1991) extend the notion of shared workspace outward from the drawing surface, capturing and conveying participants' hand gestures above the drawing surface as well. Clearboard (Ishii & Kobayashi, 1993) extends the shared workspace even further by giving users the sense that they are drawing on a plate of glass suspended between them. Participants can see their own drawing while the system creates the illusion that the other participant is drawing on the other side of the transparent surface. In this way, participants have simultaneous access to the workspace and the other participant — the seam between these two resources imposed by most other systems has been effectively erased.

Group authoring is another area that has received specialized attention. The GROVE system (Ellis, Gibbs et al., 1991) maintains a central copy of a text document being edited, providing multiple participants with independent views and allowing them all to simultaneously modify the document. Disedit (Knister & Prakash, 1990) provides similar functionality in a slightly more abstract package, allowing participants to use any editor they like to modify the shared document.

Finally, it should be emphasized that a shared electronic task representation generally plays a *supportive* role in collaborative activity — it is not the only means of communication between participants. For instance, collaborators in GROVE are connected by an audio link; participants using VideoWhiteboard were given an audio/video connection to allow them to communicate as they sketched. A recent study by McCarthy, Miles and Monk (1991) provides some preliminary insights regarding the mutually supportive relationship between shared talk and mutual access to graphical problem representations.

1.2 Evaluating Communicative Efficacy

From the survey presented in the preceding section, it is evident that a wide variety of technologies have been explored in the effort to provide computer-mediated support for distributed interaction. The type of the connection between participants provided by these systems ranges from shared textual spaces, to free-form drawing and sketching workspaces, to audio-video interfaces, to advanced virtual realities. The level of interactivity embodied in the connection varies as well, ranging from message-based systems in which the interaction is extended over hours or even days, to high-fidelity, real-time environments supported by dedicated high performance networks and special purpose input devices.

Regardless of the technologies used or the specific type of interactions supported by individual systems, all of these systems are motivated by the same underlying design goal: *to electronically support interactions among distributed participants that are functionally equivalent to copresent interactions* — the system should allow participants to manage their communicative interaction with the same ease, efficiency, and accuracy as they would if they were copresent. In other words, the basic goal of any technologically-mediated environment is to support the same *communicative efficacy* as copresent interaction. The notion of communicative efficacy provides a convenient way of referring to the extent to which a communication environment supports the communicative endeavors of participants interacting in that environment; participants in environments with a low communicative efficacy will experience more difficulty communicating and collaboratively accomplishing a task than those participants interacting in an environment with a higher communicative efficacy.

An obvious question raised by this discussion is whether any of the technologically-mediated systems for distributed interaction surveyed in the preceding section actually succeed in supporting the same communicative efficacy as copresent interaction. This issue frames two closely-related sets of research issues addressed in this dissertation:

<u>Practical Issues:</u>

- Is the communicative efficacy of technologically-mediated communication environments ever equivalent to copresent interaction?
- How can differences in communicative efficacy be related to the design decisions made in creating an environment?

- Are there fundamental limitations on the communicative efficacy supported by a technologically-mediated environment that are related to the technologies used to simulate copresence?

Methodological Issues:

- How can the notion of communicative efficacy be operationalized? What features of an environment or interactions in that environment constitute the most powerful metrics for assessing communicative efficacy?
- How can the comparative evaluation of communicative efficacy inform the design of future distributed environments?

The relationship between these two sets of research questions is apparent: the only way we can address the practical issues is by first developing a methodology for articulating, evaluating and comparing the communicative efficacy of various environments.

To establish a basis for addressing the methodological issues raised above, the following sections examine several techniques for establishing and comparing the communicative efficacy of technologically-mediated environments that have been explored in recent years. We begin by considering how the designers of existing technologically-mediated environments characterize their communicative efficacy and justify their designs. We then review recently developed empirical techniques for comparatively evaluating the performance of distributed environments. The critical examination of all of these evaluative approaches motivates the development of a more powerful technique for evaluating the performance of technologically-mediated environments to serve as the methodological foundation for the comparative analysis of communicative efficacy presented in this dissertation.

1.2.1 Evaluating the Performance of Existing Systems

The only way to justify the design of any engineered artifact and to argue its superiority over competing designs is by somehow evaluating how effectively that artifact accomplishes the task for which it was designed. That is, to what extent does the artifact satisfy the design goals that motivate the design of the artifact? For technologically-mediated communication environments, the overall design goal is to support the same communicative efficacy as copresent interaction. In this section, we consider the question of how existing systems are rationalized with respect to this goal.

In examining the literature surveyed in Section 1.2, it is evident that there has been almost no attempt at all to empirically evaluate the performance of existing systems. All accounts focus narrowly on technical aspects of the electronic environment like the physical appearance of the interface, the bandwidth of the audio and video channels, the hardware and software used, and the technical obstacles encountered in the design process. In particular, there is rarely explicit discussion of how these technical characteristics collectively contribute to the communicative efficacy of the system, much less any attempt to support such claims through empirical evaluation. At best, anecdotal accounts are offered as evidence that the system was usable and accepted by the participants it was designed to support.

At the same time, we must assume that the development of technologically-mediated communication environments has not been completely haphazard. Each new system is presumably motivated by some expectation on the part of its designers that the proposed design somehow represents an improvement on previous efforts. What this implies is that there is some non-empirical metric for characterizing the communicative efficacy of electronic environments that designers rely on to rationalize their design efforts. This tacit design rationale can be exposed by comparing the various systems produced in terms of the bandwidth of the connection between participants that they support.

The diagonal arrow in Figure 1.1 indicates the prevailing trend in the development of technologically-mediated communication environments in recent years, emphasizing that successive designs have focused on continually increasing the bandwidth of the connection between participants. For instance, the COMMUNE (Bly & Minneman, 1990) system connects users with an audio-link while they collaboratively modify a shared drawing area, while later systems like Clearboard (Ishii & Kobayashi, 1993) and Majik (Okada, Maeda, Ichikawaa, & Matsushita, 1994) provide increasingly more robust video connections between collaborating participants as well. Similarly, systems like Polyscope (Borning & Travers, 1991) , which provided only static snapshots of the conversational partners are supplanted by high-bandwidth systems like NV (Lawrence Livermore Labs, 1992) which strive to provide a more fluid, continuous video channel. Clearly, the assumption is that an audio-video environment inherently provides for better communicative efficacy than an audio-only environment which, in turn, is better than a typed-text environment. In the same way, color images are assumed to be better than black-and-white, and higher frame rates better than lower ones.

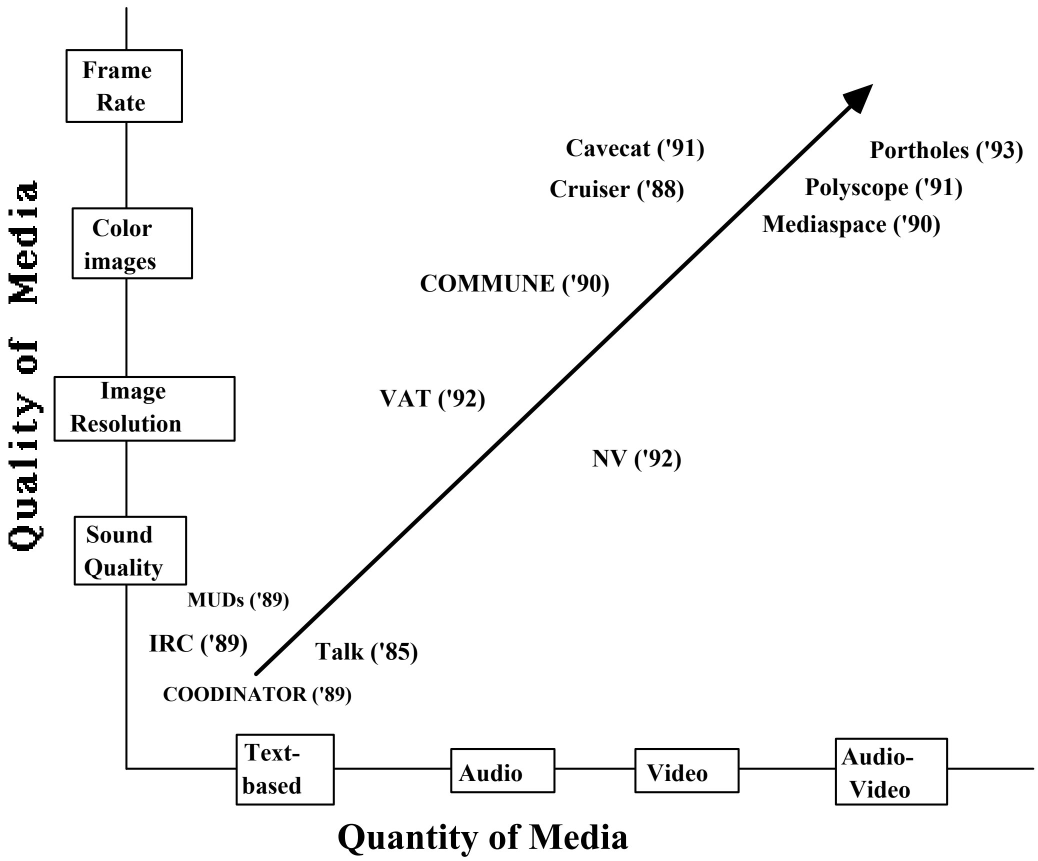


Figure 1.1: Comparison of technologically-mediated environments by the bandwidth they support. The arrow indicates the emphasis on increasingly higher bandwidth.

In general, the analysis of existing work in technologically-mediated communication reveals a pervasive underlying assumption that communicative efficacy of technologically-mediated environments is directly related to the bandwidth they support. This assumption, which we will call the *Bandwidth Assumption*, is as follows:

> The Bandwidth Assumption: The communicative efficacy of a technologically-mediated environment is determined by the bandwidth of the connection between participants that it provides. Higher bandwidths necessarily lead to higher communicative efficacy and, therefore, more robust simulations of copresent interaction.

The Bandwidth Assumption provides a trivial solution to the problem of evaluating the performance of technologically-mediated environments by establishing a heuristic relationship between easily measured physical characteristics of the environment (i.e. the bandwidth) and the extent to which the environment satisfies its design goal (i.e. communicative efficacy). It also provides a straightforward rationale for the design of future systems, placing the emphasis of such efforts squarely on further increases in bandwidth.

One reason to be suspicious of the Bandwidth Assumption as a basis for characterizing the communicative efficacy of distributed environments is obvious: it is simply an assumption. Specifically, it requires that we blindly accept that the communicative efficacy of an environment — the extent to which it supports the construction of shared understanding by interacting participants — is solely and necessarily established by the bandwidth of the connection provided by the system. By focusing on the total volume of information transmitted rather than on what information is actually used by participants as they work to maintain shared understandings of their interaction, the Bandwidth Assumption divorces the notion of communicative efficacy from the actual communicative experiences of users. While this is convenient for designers, allowing them to essentially evaluate and rationalize their designs without investing effort in empirical evaluation, the validity of this evaluative approach is clearly open to question.

In sum, the Bandwidth Assumption that motivates the design of existing technologically-mediated environments can not be considered a reliable basis for understanding the extent to which these environments succeed as simulations of copresence because it is based on the unproven assumption that higher bandwidth necessarily results in higher communicative efficacy. The notion of communicative efficacy is, by definition, a characterization of the *practical* utility of a communication environment and, therefore, cannot be directly assessed by non-empirical means. In the abscence of empirical studies that validate the Bandwidth Assumption, designing electronic environments by the Bandwidth Assumption is like designing a series of rockets based on an unproven ballistic theory, without ever testing designs to see whether they actually fly. One of the ancillary goals of this dissertation is to test the empirical validity of the Bandwidth Assumption, by evaluating and comparing the communicative efficacy of two technologically-mediated environments that support vastly different bandwidths.

1.2.2 Empirical Evaluation of Performance

A central element of the critique leveled at the Bandwidth Assumption is that it totally ignores the actual communicative experiences of participants as they interact to accomplish their collaborative endeavors in a technologically-mediated environment. Several research efforts have recently attempted to address this shortcoming by establishing empirical bases for comparing the performance of technologically-mediated communication environments. The metrics used to compare interactions can be roughly categorized into three groups: user satisfaction, quality of work, and task-activity structure. The following paragraphs briefly review each approach.

<u>User Satisfaction.</u> By far the most common technique for empirically comparing the performance of technologically-mediated environments is to rely on users' perceptions of the communicative efficacy of the system. The most direct way to expose these conceptions is to simply ask users to fill out some sort of survey (Apperley & Masoodian, 1995; Dykstra-Erickson, Rudman et al., 1995; Olson, Olson, & Meader, 1995) asking them to compare and contrast their experiences in each of several

environments being compared. A more roundabout approach is to provide users with access to several different communication environments and then track their usage patterns over time to see which environment users seem to prefer (Tang & Isaacs, 1992) . Some studies (Isaacs, Morris, Rodriguez, & Tang, 1995; Tang, Isaacs, & Rua, 1994) have combined both survey and usage data into an overall assessment of user satisfaction. After user satisfaction has been measured in some way, the results are quantified and the resulting user satisfaction ratings statistically compared between environments. Differences in communicative efficacy are implied by significant differences in user satisfaction ratings.

Quality of work. Another way to articulate differences between communication environments is to compare the quality of the work produced by interacting participants. Groups of participants working in different communication environments are given the same collaborative task to perform; the quality of the task solutions produced by each group is quantitatively evaluated and statistically compared between environments. For example, in the comparative evaluation performed by Olson et al. (1995) , groups of participants interacting in various environments were asked to design an automated post office. The resulting designs were evaluated for completeness and correctness, and the results used to statistically compare performance of the various environments. The relative communicative efficacy of environments is implied by significant differences in the quality of work.

Task-activity structure. Both the evaluation by user satisfaction and by quality of work rely on the *outcomes* of interaction as a measure of how effective the communicative interaction of participants was. Another way to express the differences between environments is by comparing the kinds of activities that participants engaged during the task-solution process. The most straightforward way to do this is to simply document the behaviors that participants engage in as they collaborate. For example, the analysis of how copresent users accomplish a collaborative design task (Tatar, 1989) was used to motivate the design of the Videodraw (Tang & Minneman, 1990) system[5], rationalizing features like multiple independent drawing tools and the ability to mark simultaneously in the electronic space. A more abstract approach to characterizing task-activity structure is to categorize the task solution activities that participants engaged in, and then to compare the amount of time invested in each activity. For example, Olson et al. (1995) identified activities like meeting management, planning and writing, digression, and summarization, and compared the amount of time spent in each and the flow of interaction from one task to another between environments.

The comparison of communicative efficacy yielded by task-activity analysis is clearly quite different from that yielded by comparing user satisfaction or quality of work. Where the two latter metrics yield an ordinal ranking of environments by communicative efficacy, task-activity analysis merely yields a nominal comparison of the structure of interaction. For example, a comparison of technologically-mediated interactions to copresent interactions by task-activity structure might reveal differences in how participants accomplish the given task, but does not naturally suggest which interactions were more effective. However, differences in communicative efficacy can be inferred from differences in the distribution of conversational effort exposed by the analysis. For instance, the observation that participants invested more time in "meeting management" in distributed interactions than when copresent might suggest that the technologically-mediated environment had a lower communicative efficacy.

1.2.3 Critique: Indirectness of Current Empirical Approaches

Each of the empirical techniques for comparing the performance of technologically-mediated environments discussed above provides some basis for inferring the communicative efficacy of the environments in which the observed interactions took place. However, all of these techniques provide only *indirect* measures of communicative efficacy in that they fail to explicitly account for the success or failure of the communicative exchanges of participants interacting within the environments. Instead, the techniques rely on the outcomes or structure of interactions as an indication of how well participants were able to communicate. For example, the evaluation of communicative efficacy based on user satisfaction turns on the assumption that perceived satisfaction directly reflects the communicative efficacy of the environment. Similarly, comparison of environments on the basis of quality of work assumes that the quality of work will necessarily suffer in environments with a low communicative efficacy.

Metaphorically speaking, using metrics like user satisfaction, quality of work, and task-activity structure as a basis for evaluating communicative efficacy is like measuring the size of a fire by the amount of smoke produced, rather than by directly investigating what it is that is burning. Specifically, three critiques can be leveled against these techniques:

1. Validity. The assumption that the metrics being measured are somehow proportional to communicative efficacy is open to question. For example, it is not clear that highly motivated participants wouldn't produce excellent work despite being hampered by the low communicative efficacy of their environment. Similarly, there may be factors entirely unrelated to the efficacy of communication that could lead to low user satisfaction ratios.

2. Accuracy. Even if we assume that the metrics measured by these techniques are directly related to communicative efficacy, there is no way of knowing how sensitive that relationship is. That is, how large do differences in communicative efficacy have to be in order to be reflected in metrics like user satisfaction and quality of work? The relevance of this question is highlighted by evidence (Olson, Olson et al., 1995) that differences in user satisfaction are not necessarily mirrored by differences in quality of work.

3. Articulation. Perhaps the greatest drawback of these techniques is that they yield no insights into what is actually going awry in the communicative interactions of participants. For example, differences in user satisfaction may imply that a difference in communicative efficacy exists, but do not reveal the communicative troubles experienced by users that are presumably the root cause of their dissatisfaction. That is, measuring the amount of smoke produced by a fire does not give any insight into what it is that is actually burning and, consequently, provides no basis for understanding what caused the fire in the first place.

In sum, existing empirical techniques all rely on metrics that measure communicative efficacy indirectly, by comparing the outcomes or abstract structure of interaction. In particular, they do not directly expose the ways in which a communication environment supports or impedes the efforts of participants to reach a shared understanding of their interaction.

1.3 Summary and Discussion: Evaluating Communicative Efficacy

The purpose of the chapter has been to motivate the research presented in this dissertation by highlighting serious shortcomings in the way in which technologically-mediated environments are designed and evaluated. The basic premise for the discussion presented in this chapter is that the goal of any technologically-mediated environment is to allow widely distributed participants to communicate and accomplish collaborative tasks with the same ease, accuracy and efficiency as if they were copresent. The notion of communicative efficacy was introduced as a way of reifying the level of support that an environment provides for the communicative endeavors of participants. Participants working in environments with a high communicative efficacy will have less difficulty maintaining a shared understanding of their collaborative activities, while interactions in environments with a low communicative efficacy will be marred by misinterpretation, confusion, and a general failure of participants to work together towards a mutually satisfactory solution to the tasks they are engaged in.

The fundamental goal of any technologically-mediated communication environment, therefore, is to provide the same communicative efficacy as copresent interaction. This observation leads directly to the research issue addressed in this dissertation: Do existing technologically-mediated environments actually provide robust simulations of copresence, providing the same communicative efficacy as copresent interaction?

The answer to this question clearly depends on how the communicative efficacy of an environment is characterized and evaluated. An analysis of existing work reveals a general failure to formally evaluate the performance of technologically-mediated environments by empirically comparing interactions in those environments to copresent interaction. Instead, designers have tacitly relied on abstract technical metrics for characterizing the performance of the communication environment. Specifically, there is an overall assumption that the communicative efficacy of a technologically-mediated environment is directly related to the bandwidth of the connection between participants it provides. This single-minded devotion to technical issues has been both a curse and a blessing. On the one hand, it has motivated tremendous technical achievements in network management, data compression, and a plethora of creative software applications to go with them. As a result, designers of distributed communication environments have far more implementional options than just a few years ago. On the other hand, the failure to empirically evaluate the performance of technologically-mediated environments represents a fundamental failure to justify the design of electronic environments in terms of the overall design goal of supporting the same communicative efficacy as copresent interaction. Without a detailed analysis of how real participants are actually able to use the electronic simulation of copresence afforded by an environment to accomplish their communicative goals, there is the clear danger that the entire technical thrust will miss the mark.

In response to this criticism, several empirical techniques for comparing the performance of technologically mediated environments have been developed in recent years. An analysis of these techniques reveals that they rely on indirect metrics to infer the communicative efficacy of communication environments. This makes these techniques essentially identical to the Bandwidth Assumption in that both rely on abstract characteristics to infer communicative efficacy — where the Bandwidth Assumption infers the communicative efficacy based on technical characteristics, existing empirical approaches infer communicative efficacy based on the outcomes or structure of interaction. In particular, both approaches fail to directly examine the very source of communicative efficacy, namely, the communicative interaction of participants itself. These observations are graphically summarized in Figure 1.2.

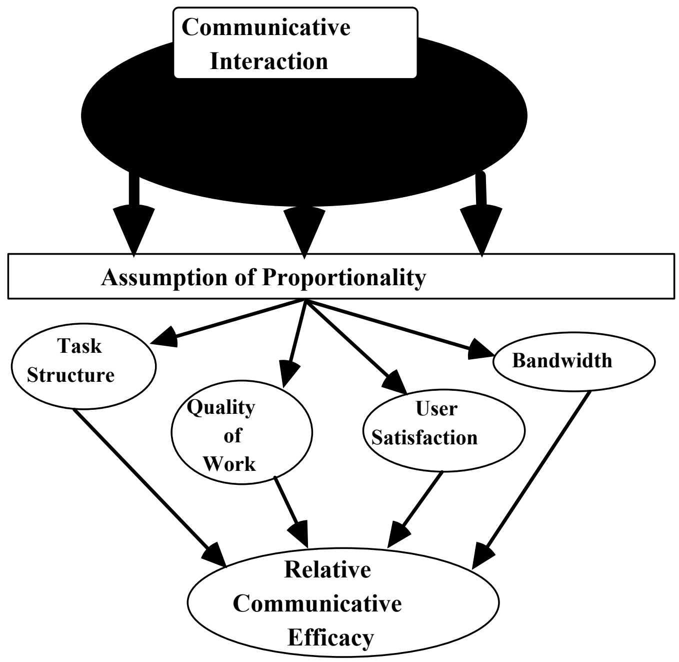


Figure 1.2: Existing approaches to evaluating communicative efficacy all fail to directly examine the communicative interaction of participants.

Figure 1.2 clearly illustrates how both the Bandwidth Assumption and empirical approaches evaluate communicative efficacy *indirectly*, by comparing either abstract features of the environment or overall outcomes of interactions in those environments to infer the extent to which participants are able to communicate successfully in those environments. As a result, none of these evaluative approaches provides a suitable methodological foundation for the comparative analysis of communicative efficacy undertaken in this dissertation. However, the deficiencies revealed by the analysis of existing work provide a strong foundation for developing a more powerful evaluative methodology for characterizing and comparing the communicative efficacy of technologically-mediated environments. Specifically, a useful methodology must meet the following criteria.

1. Empiricism. Any evaluation of communicative efficacy must be based on the analysis of real participants engaged in real tasks. The notion of communicative efficacy is inherently pragmatic and is inherently grounded in the communicative experiences of real users.

2. Directness. The only way to measure communicative efficacy is by focusing the analysis *directly* on the communicative interactions that take place in that environment, examining the ways in which these interactions succeed or fail at establishing shared understanding.

3. Explanatory Power. A viable methodology must concisely characterize the communicative troubles experienced by participants, rather than merely exposing overall differences in communicative efficacy. This concise articulation of *what* is going wrong during communicative interactions is vital for explaining *why* it is going wrong and, ultimately, for understanding how deficiencies in communicative efficacy are related to the physical characteristics of the design.

The central theme of these criteria is that the comparative evaluation of communication environments must be based on characterizing the extent to which communicative interaction results in shared understanding. In this way, the communicative efficacy of a communication environment is defined by its *epistemic* performance — the extent to which it actually supports the collaborative construction of shared meaning by interacting participants — rather than its physical characteristics. This observation exposes a fundamental difference between the evaluation of technologically-mediated communication environments and the evaluation of almost all other engineered artifacts. For example, a toaster can be straightforwardly evaluated by the physical condition of the toast it produces; the performance of a space shuttle can be evaluated by examining its physical behavior to see if it matches the expectations of designers. By contrast, there exists no direct physical means of examining communicative efficacy — the aim of establishing "shared understanding" represents an epistemological goal rather than a physical one. Accordingly, any methodology for assessing communicative efficacy must be based on a solid epistemological foundation that articulates what it means to know and understand, how shared understanding arises through communicative interaction, and how to conceptualize communicative trouble.

1.4 Overview of Solution

The methodology used to explore the communicative efficacy of technologically-mediated environments in this dissertation is based on an epistemological foundation recently developed by social scientists known as Situated Action. The basic premise of Situated Action is that the significance of action arises dynamically and uniquely in the interplay between an observer's past experience and the contingencies of the local context of interpretation; communication is characterized as a collaborative construction of the significance of mutually available experience, in which each participant continuously makes available evidence of his or her interpretive orientation, while simultaneously interpreting the communicative displays of others. In this way, shared meaning is not a final result of communicative interaction, but is fluidly negotiated throughout the interaction.

The closely related methodologies of Conversation Analysis and Interaction Analysis have been developed by ethnomethodologists specifically to expose the way in which shared understanding arises though communicative interaction, by documenting the conversational regularities that interacting participants rely on to organize their contributions to the interaction and maintain shared interpretations of mutually available events. Importantly, the way in which these regularities become apparent to the analyst is when they are somehow violated, resulting in communicative confusion of some sort. This notion of *communicative breakdown* provides the cornerstone of the evaluative methodology developed in this dissertation, which we will call Breakdown Analysis. Specifically, Breakdown Analysis is based on the following ideas:

1. Communicative breakdown directly embodies the notion of communicative efficacy. In particular, the communicative efficacy of interaction is reflected in the amount of communicative breakdown experienced by participants; the lower the incidence of breakdown, the higher the communicative efficacy of the interaction.

2. Interaction analysis constitutes a strong methodological foundation for empirically evaluating communicative efficacy, providing powerful analytic techniques for exposing and characterizing the communicative breakdowns in naturally-occurring interactions. Though Interaction Analysis is by nature a purely documentary technique, it can be modified and extended to create a viable methodology for stochastically comparing the communicative efficacy of interaction in different communication environments.

Briefly, Breakdown Analysis is based on the intertwining of the qualitative techniques afforded by Interaction Analysis with the quantitative techniques of traditional scientific investigation to yield a powerful analytic tool for empirically comparing the performance of two or more communication environments. The methodology consists of three phases, which progressively refine our understanding of the differences in communicative efficacy that exist between the environments being compared:

<u>Phase One: Recognizing Breakdown.</u> After an initial data collection effort, during which the interactions of pairs of participants in each communication environment are captured on videotape and transcribed, the first phase of the analysis applies the qualitative techniques of Interaction Analysis to articulate consistent patterns of

communicative breakdown and to establish strong, consistent evidentiary criteria for recognizing breakdowns in each category.

Phase Two: Exposing Differences in Communicative Efficacy. In the second, quantitative phase of the analysis, the criteria developed in phase one are applied to expose all breakdowns in each category that occurred over the course of each interaction. The number of breakdowns documented in each category is used as a direct metric for communicative efficacy; a statistical comparison of the total amount of breakdown between environments is used to expose significant differences in communicative efficacy.

Phase Three: Rationalizing Differences. In the final phase of the analysis, the differences in communicative efficacy exposed in phase two are used to motivate and focus a second qualitative analysis aimed at explaining why those differences exist. By establishing causal relationships between certain physical characteristics of an environment and the higher incidence of communicative breakdowns observed in that environment, this analysis establishes a strong basis for future redesign.

Clearly, the methodology of Breakdown Analysis satisfies the methodological criteria laid out earlier in Section 1.3: it is empirical, since it is based on the actual communicative experiences of participants; it is direct, in that it focuses analytic attention specifically on the moment-by-moment communicative behaviors of participants; and it supports rationalization of differences in communicative efficacy by revealing how breakdowns are related to the design of a communication environment.

Breakdown Analysis provides a firm methodological foundation for addressing the research issues raised in this chapter, by exploring the functional differences between copresent and technologically-mediated interaction. Specifically, Breakdown Analysis was used to compare the communicative efficacy of interaction in three very different communication environments: copresent interaction, in which participants were seated side-by-side; audio-only interaction, in which participants were in separate rooms, communicating via an audio-link; and audio-video interaction, in which participants were, again, in separate rooms, but now had both an audio and a video connection. In all three scenarios, participants had shared access to an electronic workspace, using a simulator running in the shared workspace to collaboratively accomplish a series of non-trivial tasks.

These three environments — copresent, audio-only, and audio-video — were selected for comparison for several reasons. Most importantly, they canonically represent the basic media choices that are currently available to designers of modern technologically-mediated environments. In this way, the insights yielded by this analysis should be relevant to a broad range of design contexts. A second reason for including both an audio-only and an audio-video environment in the analysis is that the comparison of relative communicative efficacy of these two environments explicitly tests the Bandwidth Assumption, which tacitly underlies (see Section 1.2.1) many current design efforts. If the Bandwidth Assumption is valid, the analysis should reveal that audio-only interactions have a significantly lower communicative efficacy than audio-video interactions, since the latter environment clearly provides a higher bandwidth connection between participants.

The remaining chapters of this dissertation elaborate on the overview presented in this section, and then present the results of applying Breakdown Analysis to the three communication environments described above. Chapter II establishes the epistemological and methodological foundations of Breakdown Analysis, presenting an in-depth discussion of Situated Action and contrasting it with traditional Representationalist models of cognition. The methodologies of Conversation and

Interaction Analysis are then reviewed, providing a strong basis for understanding their relationship to Breakdown Analysis. Chapter III lays the groundwork for the comparative study of copresent and distributed interaction, formally introducing the methodology of Breakdown Analysis (i.e. the analytic tool), and describing the three communication environments that were compared in more detail. The next three chapters then present, respectively, the results of each of the three phases of the Breakdown Analysis: Chapter IV presents the results of the initial qualitative study, detailing the patterns of breakdown that were identified in the analysis and how they were operationalized; Chapter V presents the results of the second, quantitative phase of the analysis, statistically comparing the frequency of breakdowns documented in each environment, and drawing conclusions about differences in communicative efficacy; in Chapter VI, these differences are used to drive a focused qualitative investigation aimed at rationalizing the observed differences in communicative efficacy in terms of resource constraints imposed by technologically-mediated environments. Finally, Chapter VII summarizes results and discusses the implications of this analysis for the design of technologically-mediated environments in general.

1.5 Notes

[1] Some users may also be accorded the privilege to "build" onto the virtual space, defining new rooms. In this sense, the MUD is designed by no one user; this has proven to have considerable appeal (Curtis, 1992)
.
[2] This assumption regarding the communicative value of video connections is examined in more detail in section 3.1.
[3] The majority of task-oriented systems are reviewed in section 2.2. As mentioned earlier, the decision as to which section a system belongs to was made by judging whether its *primary* focus was on personal interaction or task-oriented interaction.
[4] In their current nascent implementations, neither of these systems are actually computer-based, relying only on cameras and displays connected by conventional analog connections.
[5] The fact that the design of Videodraw is motivated at least in part by empirical observations makes it a rare exception to the overall tendency to rationalize design purely in terms of bandwidth. Despite these promising beginnings, however, the success of Videodraw — whether its simulation of empirically observed features of task activity actually worked — was never empirically evaluated.

CHAPTER II
BACKGROUND

In the preceding chapter, it was suggested that the only way to assess the communicative efficacy of interaction is to directly account for the way in which shared understanding arises — or fails to arise — from the communicative behaviors of conversants. A critical examination of existing approaches revealed that these approaches all measure communicative efficacy indirectly, relying on a variety of abstract characteristics of the environment or interactions in the environment to infer how effectively participants were able to communicate during an interaction. This observation motivates the development of a novel methodology that directly analyzes the communicative process by which participants arrive at a shared understanding of mutually available events, somehow documenting the extent to which this process succeeds and fails over the course of an interaction.

Clearly, developing a methodology for assessing communicative efficacy centered around differences in the "understanding" of participants requires careful consideration of fundamental philosophical questions related to human cognition and communication. In order to comprehend the notion of shared understanding, we must first articulate how an individual finds meaning in experience and how this interpretive process is related to rational action. Only then can we begin to consider how meaning might be "shared" between conversational participants, and how failures to establish shared meaning might be manifested and reliably detected in the communicative behaviors of participants as a way of measuring the efficacy of communicative interaction. Specifically, two issues must be addressed:

1. How does communication work in principle? What does it mean for a person to “know” or “understand” something, how does communicative interaction make that knowledge available to a conversational partner, and how are we to conceive of failures in this communicative process?

2. How can we empirically measure the communicative efficacy of interactions by documenting the success or failure of the communication process?

The first issue emphasizes the need to ground the development of any evaluative methodology on a firm epistemological foundation. Any attempt to analyze the extent to which communicative interaction results in shared understanding must be based on a strong conception of what it means to “know,” and how communicative interaction makes that knowledge available to others. The second issue is more pragmatic, raising the question of how intangible, internalized mental processes and events are evidenced in and can be inferred from an examination of observable communicative behaviors.

The goal of this chapter is to provide a brief review of the bodies of work that the evaluative technique used in this research draws on to address each of these issues. The following section establishes the epistemological foundations of the dissertation by introducing the epistemology of Situated Action, and contrasting it with more traditional Representationalist conceptions of cognition and rational action. Based on this discussion, communicative interaction is characterized as a collaborative, evidentiary process in which participants continuously construct the significance of ongoing action based on the contextual interpretation of each other’s communicative displays. Section 2.3 then introduces the closely-related methodologies of Conversation and Interaction Analysis, which have been developed to analyze and document the way in which shared interpretations of action are constructed and maintained by interacting participants. Because these methodologies explicitly reveal the ways in which communicative interaction fails, or breaks down, they constitute strong methodological foundations for evaluating the communicative efficacy of interaction.

2.1 Epistemology: What is Communication?

Communication is a pervasive component of almost all organized human endeavors, from landing rockets on the moon, to teaching and learning, to ordering a cheeseburger at a fast-food restaurant. With such a wide range of communicative contexts, it is difficult to decide on a framework for understanding how communication works — how interacting participants arrive at shared interpretations of a given situation or, alternatively, how they might *fail* to arrive at the same interpretation. One way to simplify the discussion is by abstracting away from the infinite differences between communicative scenarios, and recognizing that all communicative interactions share a common underlying goal: one or both participants are trying to somehow change their partner's interpretation of mutually available events. In other words, the goal of all communication is to *change* a partner's knowledge of the world, by making one's own interpretation of a given situation known to him or her.

This insight provides us with the analytic leverage we need to compare and contrast two very different epistemological frameworks that have been proposed for understanding what it means to know, and how knowledge is related to rational action: Representationalism and Situated Action. These two epistemological frameworks differ profoundly in their characterization of *what* it is that is changed during communicative episodes and *how* that change is brought about. How one answers these epistemological questions — how meaning arises and how it is communicated — fundamentally shapes one's approach to interpreting, describing, and rationalizing the efficacy of communicative interaction.

Though the research presented in the following chapters is ultimately based on the epistemology of Situated Action, there is a rhetorical advantage to beginning the discussion with a review of the Representationalist epistemology. Because Representationalism is based on finite symbolic representations of meaning and mechanistic conceptions of human reasoning, it is much easier to describe and comprehend than the dynamic, ephemeral characterization of meaning and its relationship to rational action promoted by Situated Action. Accordingly, the most effective way to introduce the notion of Situated Action is by directly contrasting it with traditional Representationalist conceptions of cognition and communication.

2.1.1 Traditional Conceptions of Cognition: Representationalism

Representationalism has been identified (Doerry, 1994; Newell, 1980; Suchman, 1987; Winograd & Flores, 1986) as the epistemological basis for nearly all modern scientific disciplines related to the study of human cognition, including cognitive science, psychology, and artificial intelligence. The central tenet of Representationalism is that we carry inside our heads symbolic models, or *representations*, of the physical world and its behavior as well as of our intentions, goals, and beliefs with respect to the world, and the actions that we can perform (i.e. plans) to achieve certain goals; these symbolic models serve as the basis for all reasoning and action that we perform.

As indicated in Figure 2.1, rational action under the Representationalist epistemology is strictly goal-directed, or *intentional*, based on the manipulation of symbolic models of the current context to produce detailed plans for accomplishing specific goals— intelligent behavior is the result of implementing these plans.

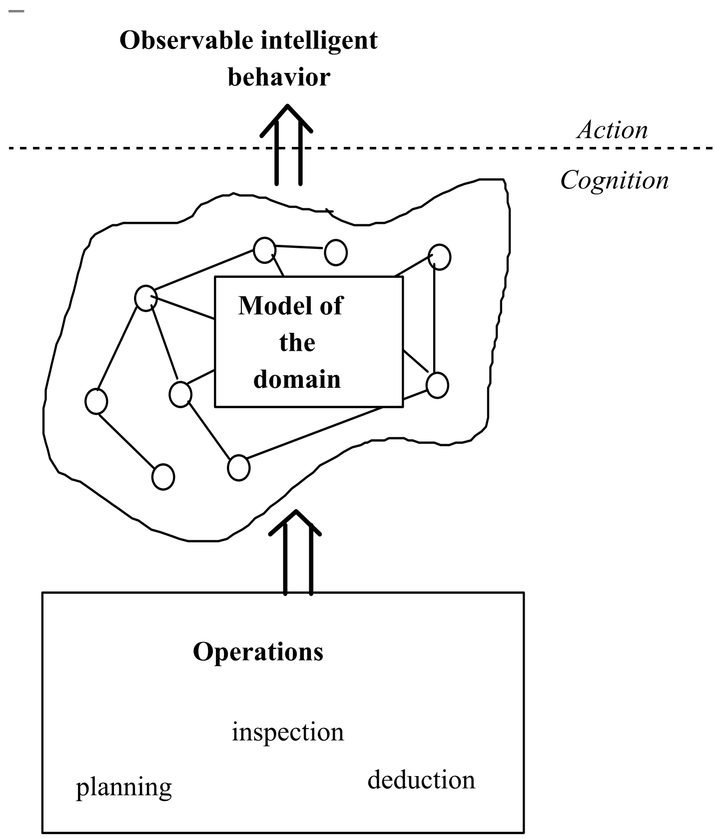


Figure 2.1: Representationalist model of rational action.

To account for events in an ever-changing world, the Representationalist epistemology asserts that the symbolic reasoner continuously updates its symbolic representations of the world by accepting any number of external inputs, which are presumed to be connected directly to various sensory devices (e.g. eyeballs, video cameras, etc.). However, the number of such inputs, though perhaps quite large, must always be finite and specified in advance. More importantly, the significance ascribed to a perceived event — how it may influence the outcome of reasoning — is deterministically and permanently defined at the time it is perceived by the symbolic processes used to integrate that event into the overall symbolic structure. This observation leads to the heart of the Representationalist paradigm, namely, the assumption that the significance of action can be defined succinctly, monotonically, and independently of a context of use.

2.1.1.1 Communication as Transfer of Symbolic Models

Since knowing is reduced to symbolic representation under the Representationalist epistemology, "shared understanding" is defined by a mental condition in which two participants have identical symbolic representations (i.e. mental models) of the state of the world and the deterministic processes that govern its behavior. The way in which this condition arises through communicative interaction with other intelligent agents[1] is straightforward: communication is characterized as the transfer of symbolic structures from one participant to another, as illustrated in Figure 2.2.

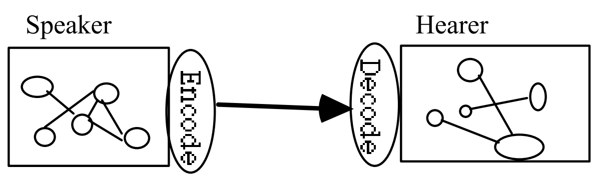


Figure 2.2: Communication in a Representationalist world

In this "conduit" (Reddy, 1979) model of communication, the symbolic knowledge of the speaker is somehow encoded into a form suited to the particular medium of communication, and is then transferred to the hearer, who decodes the information to arrive at the same (symbolic) knowledge. Because of the deterministic correspondence between symbolic forms and their significance postulated by the Representationalist epistemology, the only way that communication can fail is through an error in the transfer of these symbolic forms caused by some flaw or inadequacy in the medium.

Transfer of knowledge from outside sources is not the only way in which knowledge grows under the Representationalist paradigm; new knowledge can also be derived from existing knowledge by symbolic manipulation, in the form of logical deduction and symbolic abstraction (Laird, Newell, & Rosenbloom, 1987) . That is, new symbolic structures (and thereby novel meaning) can be derived from existing ones using certain deterministic heuristics designed to preserve the semantics of the symbolic system.

A fundamental problem with the Representationalist conception of human cognition is that, because of the essentially *static* semantics of symbolic representation, it is difficult to account for the flexible fashion in which humans are able to respond to the unpredictable contingencies of life in the real world. The fact that all knowledge is symbolic in form, embodied in internalized syntactic representations, requires that a given symbol structure must[2] inevitably have a fixed, finite meaning, which is determined at the time the symbolic form is created. Specifically, the significance of action is determined based on context-independent heuristics that deterministically define the meaning of specific behaviors. Because the significance of experience is permanently established at the time the events in question are perceived, reinterpretation of those events in light of future experience is ruled out. In the next section, we will consider a radically different epistemological foundation that avoids these shortcomings by avoiding the commitment to symbolic representation as a prerequisite to rational action.

2.1.2 Ideas from Ethnomethodology: Situated Action

The emergence of Ethnomethodology as a distinct approach within Sociology marks a breaking away from rigid Representationalist conceptions of knowing and communicating. Originally, Ethnomethodology was developed (Garfinkel, 1967) as an alternative to the “voluntaristic” theory of action (Parsons, 1937) , which holds that common values/norms (internalized during socialization) influence and motivate human action. By suggesting that action is generated by reference to context-independent knowledge, the voluntaristic theory is clearly Representationalist in nature. Specifically, the universal norms posited by the voluntaristic theory can be seen as predefined "plans" for action, which are implemented by the human actor in order to accomplish some goal. Drawing on the phenomenological writings of Schutz, Husserl, Gurwitsch, Merleau-Ponty, and Heidegger, Garfinkel stressed the knowledgeability of actors and how they use common-sense practices/procedures to produce, analyze and make sense of one another's actions and their local or *situated* circumstances. A central feature of this view is the notion that knowing and acting are reflexively intertwined; it is through concerted action in the setting of practical activity that participants create and recreate the intelligibility and "facticity" of their social *situated* world and the activities in which they are engaged.

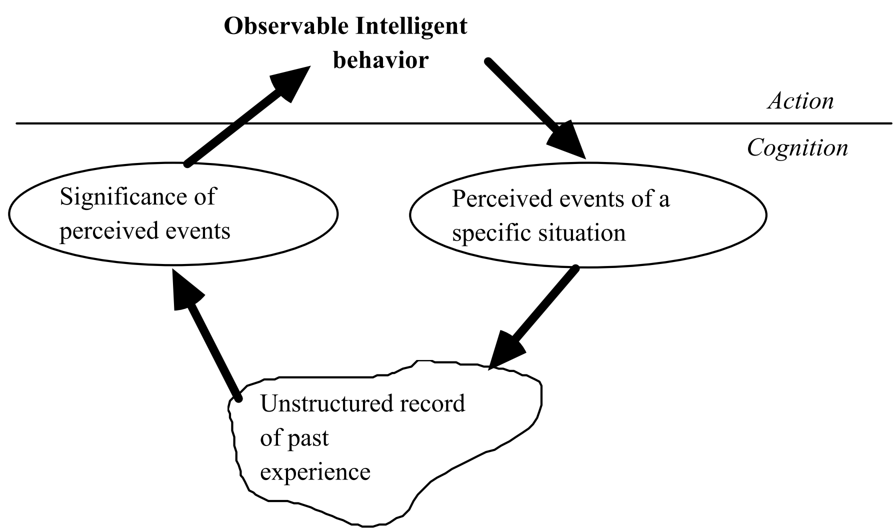


Figure 2.3: The reflexive relationship between knowing and acting posited by Situated Action.

As implied by Figure 2.3, the epistemology of Situated Action denies the existence of meaning outside of a specific context of action. Rather, knowledge under the epistemology of Situated Action is conceived of as an unstructured record of experience with no inherent semantics at all. This record of experience (Suchman, 1987) serves as an *interpretive resource* for dynamically constructing the significance of events we perceive in the world; meaning (i.e. knowing) arises in the dynamic interaction of this unarticulated, amorphous record of past experience with the particulars of the current context. It is this locally constructed interpretation of action that serves as the basis for and motivates our rational actions. Importantly, these actions themselves constitute events in the local context, inevitably changing that context and thereby influencing the significance of current and previous events. In this way, knowledge arises only in the context of ongoing, situated action; the significance of action evolves fluidly as events unfold and can not be distilled out of that context and captured in a static symbolic form.

2.1.2.1 Communication as Collaborative Interpretation

Under Situated Action, communication is characterized as a seamless extension of the dynamic interpretive process by which individuals construct the significance of action. Communication is viewed as the collaborative construction of the significance of mutually available events, in which both participants continuously make available evidence of their interpretation of ongoing action — including the communicative actions of their conversational partner — while at the same time examining the behavior of others to infer their interpretation of the evolving interaction.

Figure 2.4 illustrates this dialectic, evidentiary process of finding the meaning in action based on mutually available interpretive resources; the empty ovals in the figure emphasize that the communicative resources that may be relevant in shaping the evolving interpretation of action can not be finitely enumerated or specified in advance.

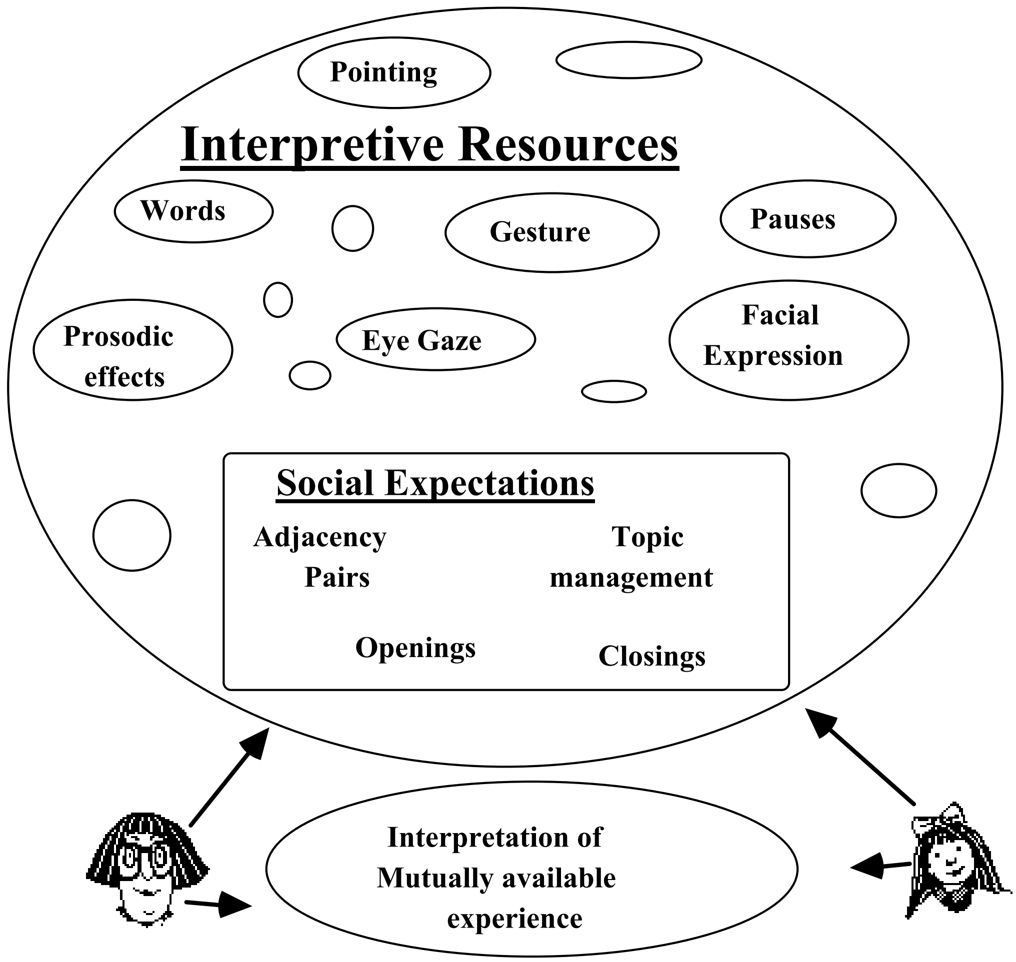


Figure 2.4: Communication as the collaborative construction of shared interpretation of action.

Under Situated Action, symbols and symbolic structures (e.g. words, diagrams, icons, etc.) have no intrinsic epistemic significance at all; they are simply linguistic tools used to rationalize and objectify action retrospectively. That is, symbolic representations play a purely descriptive role, applied retrospectively to objectify perceived action and display evidence of one's interpretation of its significance. This notion is illustrated in Figure 2.5.

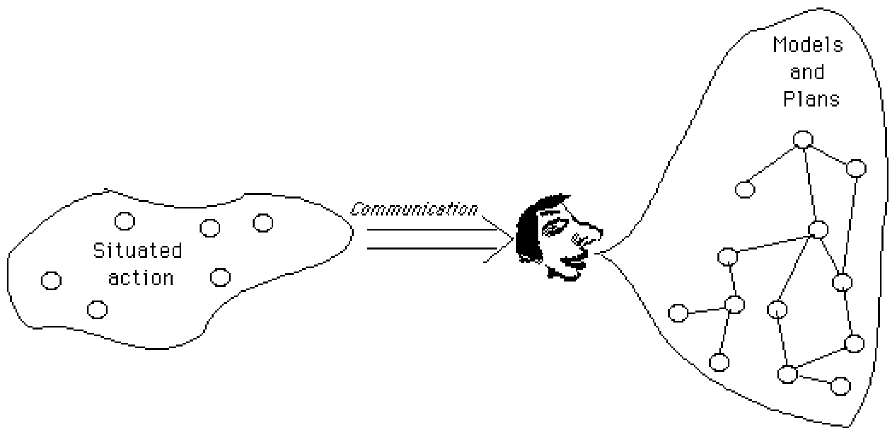


Figure 2.5: Symbolic structures as descriptive tools for objectifying action.

Clearly, this perspective on the role of symbolic representation in human cognition is directly antithetical to the perspective promoted by the Representationalist model depicted earlier in Figure 2.1: Instead of serving as knowledge models that form the *generative basis* for action, symbolic representations serve as *descriptive tools* used retrospectively to communicate about action. An important consequence of this reversal is that symbolic representations have no inherent significance whatsoever; the significance of any symbolic representation (including natural language) is constructed locally and collaboratively by participants, with respect to the unique contingencies of the immediate context of action.

Of course, humans do construct plans for future action far in advance. Situated Action allows for such plans, but only as high level organizational mechanisms used, again, as resources for objectifying and communicating about future intentions. In particular, such plans do not in any concrete way constrain the specific actions that eventually result.

2.1.3 Summary: Situation Action and Representationalism

In sum, Representationalism and the Situated Action differ fundamentally in how they conceive of knowledge, its relationship to rational action, and the way it is made available to others through communicative interaction. Where Representationalism posits symbolic manipulation as the fundamental basis for action and communication, the Situated View begins with situated action and posits language and other symbolic representations as linguistic tools used in the inherently social communicative enterprise of constructing shared interpretations of action. The key points of the two epistemologies are contrasted as follows:

Representationalism

- The significance of experience can be deterministically established and finitely represented in symbolic structures. In this way, knowing is independent of any particular context of action.
- Establishing "shared meaning" is a matter of arranging for both parties to have the same symbolic knowledge model.
- Because symbolic representations are semantically unambiguous, there is no need to negotiate over the meaning of a symbolic structure; the structure *is* the meaning.
- Communication is a matter of correctly transferring appropriate symbolic knowledge.

Situated Action

- Meaning can not exist independently of context.
- Meaning is continuously and uniquely constructed in the dynamic interaction of past experience and the contingencies of the current context of action.

- Symbolic representations (including language) have no inherent meaning at all and play no role in generating rational action. Rather, they serve as an important resource for communicating about the significance of action, by objectifying experience and thereby displaying evidence of one's interpretation of experience to a conversational partner.
- Communication is characterized as an evidentiary, interpretive process by which participants collaboratively construct shared interpretations of action.

Clearly, each of these epistemologies promotes a radically different conception of the notion of shared understanding and how it arises from communicative interaction; the way in which we conceive of communicative efficacy and how to go about evaluating and comparing the communicative efficacy of interactions is crucially dependent on the epistemological foundations that inform the analysis. The following section establishes Situated Action as the epistemological foundation for this research and concisely articulates the notion of communicative efficacy arising from this commitment.

2.2 Situated Action as a Basis for Evaluating Communicative Efficacy

For much of the history of modern scientific thought, Representationalism has served as the undisputed basis for understanding and modeling human cognition and continues as the epistemological cornerstone of nearly all cognitively-oriented disciplines. At the same time, a growing body of evidence suggests that Situated Action may provide a more flexible and powerful basis for understanding intelligent behavior. A number of convincing philosophical arguments have been advanced (Suchman, 1987; Winograd & Flores, 1986) which suggest that the finite, static semantics intrinsic to all symbolic representations make them fundamentally unable to account for the fluid, dynamic fashion in which humans are able to adapt to the infinite and unpredictable contingencies of an ever-changing world. Because Representationalism equates knowledge with symbolic forms, the meaning of such structures must be inherent in the symbols themselves — a given symbolic structure has a fixed, finite meaning. This characteristic of symbolic representation has two important consequences. First, the significance of perceived events is inherently constrained by the symbolic processes that encode perceptual stimuli into symbolic forms — though the range of possible interpretations that can be derived for a given situation may be large, it is ultimately finite and predetermined by the symbolic processes presumed to be embodied in whatever perceptual mechanisms are available to the entity. Second, the significance of perceived events is invariably fixed at the moment of symbolic representation, ruling out reinterpretation of the meaning of the event in light of future experience.

To see how these two features of the Representationalist epistemology constrain the interpretation of experience, consider a situation in which a person approaches another and says “hello”. Under the Representationalist epistemology, the only way to establish the significance of this utterance is to characterize it as one step in a plan being executed by the speaker, which ultimately connects the communicative behavior to a specific goal. Clearly, the plan and goal the interpretive mechanism posits for the speaker will be constrained by certain contextual features: if the speaker occurs in a bar, the speaker may have romantic intentions; if the speaker is a salesperson, the speaker may be initiating a sales pitch, and so on. However, the range of possibilities is predefined by the symbolic interpretive processes that infer the speaker’s plan. More specifically, the contextual features that are “relevant” (e.g. place of occurrence, who is the speaker) to constructing the significance of action are finite and specified in advance; the symbolic system can not take into account contextual features that are not represented in its internal model of the world. Furthermore, once the significance of the speaker’s actions has been determined and symbolically represented, it can not be reassessed in light of future events. For example, though it initially seemed the speaker was initiating a sales pitch, it may later turn out that the speaker is a forgotten friend; in this light, the opening “hello” takes on a very different significance as an opening to a friendly conversation.

In short, the inherently static nature of symbolic representation makes it difficult for the Representationalist model of human cognition and communication to account for the dynamic way in which the significance of action arises with respect to its unique context of occurrence. This is particularly true of communicative interaction: the significance of a given communicative behavior is intimately dependent on the unique particulars of the interaction (e.g. history of the interaction, social relationship of conversants, etc.)— it is hard to see how such fluid contextual dependencies can be reconciled with the fixed and finite semantics of symbolic representation and the mechanistic conception of communicative interaction as the deterministic, unidirectional transfer of symbolic forms associated with the Representationalism.

Perhaps some of the most compelling evidence in favor of Situated Action comes from empirical studies of human learning, which essentially represent pragmatic investigations of how humans are able to communicate most effectively, since establishing shared understanding is the ultimate goal of all learning interactions. A review of the literature in the educational sciences, for instance, reveals that a slow metamorphosis is taking place in education (Egan, 1989; Hamm & Adams, 1992; Slavin, 1983) , moving away from traditional lecture-based approaches, which are clearly based on Representationalist conceptions of communication as a unidirectional transmission of knowledge, to “collaborative learning” approaches which emphasize interactions between knowers and learners and between learners themselves as the most important resources for developing robust comprehension of new material. Studies of informal learning contexts like apprenticeship (Lave & Wenger, 1991) have also emphasized the intimately contextual, collaborative way in which participants who initially differ greatly in their level of comprehension eventually arrive at shared understanding.

In light of these philosophical and empirical arguments in support of Situated Action, a vehement debate has recently ensued (Agre, 1993; Greeno, 1993; Suchman, 1993; Vera & Simon, 1993) over whether Situated Action or Representationalism constitutes the appropriate basis for conceiving of human cognition, or whether they are truly distinct epistemologies at all[3]. This philosophical debate is unlikely to ever be definitively resolved since, in the final analysis, human cognition is an ephemeral epistemic phenomenon that can never be directly examined. Certainly it is beyond the scope of this dissertation to provide a detailed review of the arguments in favor of Situated Action, much less to attempt to significantly advance the debate. Accordingly, we merely state that, based on a careful evaluation of the evidence mentioned above, Situated Action was selected as the epistemological foundation for the comparative analysis presented in this dissertation since it appears to provide the strongest basis for understanding how shared understanding arises in naturally-occurring communicative interactions. In particular, the fact that Situated Action essentially views language as a *tool* for locally and collaboratively constructing meaning rather than as a linguistic *conveyance* for mechanistically transferring meaning seems very natural in light of our everyday experiences as language users.

In addition to philosophical and empirical evidence, there are strong practical reasons for adopting Situated Action as a basis for this research. Because it characterizes communication as the collaborative construction of shared interpretations of action, Situated Action strongly supports an empirical approach to comparatively evaluating the communicative efficacy of interaction. By contrast, the Representationalist conception of communication as a deterministic, unidirectional transmission of symbolic knowledge trivializes the complexity of communicative interaction. In particular, the conduit metaphor (see Figure 2.2) that serves as the basis for understanding communicative interaction under the Representationalist epistemology implies that communicative efficacy is not really related to the details of participants' communicative interaction at all, but is dependent *solely on the medium* that exists between interacting participants. That is, the only way in which communicative interaction can fail under the conduit metaphor is through some deficiency in the medium, since the symbolic processes by which knowledge is encoded and decoded are semantically unambiguous; evaluation of communicative efficacy is reduced to evaluating the quality of the medium. Indeed, by replacing the term "medium" with "bandwidth", it is obvious that the Bandwidth Assumption revealed in Chapter I as the design rationale tacitly used to motivate the development of current technologically-mediated environments represents a direct embodiment of the conduit metaphor.

By contrast, Situated Action locates the efficacy of communicative interaction directly in the situated, evidentiary process by which participants collaboratively construct and maintain shared interpretations of action. In other words, the epistemology of Situated Action both motivates and justifies the empirical analysis of participants' moment-by-moment communicative behaviors to reveal the extent to which communication results in shared understanding.

2.2.1 Communicative Efficacy Under Situated Action

The commitment to Situated Action as an epistemological basis for analyzing the efficacy of communicative interaction makes it possible to define more concisely the notion of communicative efficacy:

> Communicative Efficacy refers to the extent to which interacting participants are able to establish and sustain shared interpretations of their communicative interaction as it evolves. Interactions with low communicative efficacy are distinguished from those with high communicative efficacy by a greater prevalence of divergent interpretations of action, manifested as communicative confusion or *breakdown.*

This concise articulation of what it means for communicative interaction to be effective provides a strong theoretical foundation for comparing the communicative efficacy of copresent and technologically-mediated interaction. Specifically, it suggests that the appropriate way to establish the relative communicative efficacy of interactions is to somehow compare the amount of communicative breakdown that occurs in those interactions. The following section establishes the practical foundations for such an analysis by introducing two closely-related methodologies which can be used to expose the communicative breakdowns present in naturally-occurring interactions.

2.3 Methodologies: Exposing Communicative Breakdown

While the notion of communicative breakdown provides a direct metric for characterizing the communicative efficacy of interaction, it raises the obvious question of how one might go about identifying such breakdowns in real interactions. In this section, two closely-related methodologies are introduced that are designed to document how people achieve mutual intelligibility in their everyday communicative interactions, by revealing certain conversational regularities that interacting participants rely on to shape interpretation of each other's communicative behaviors and organize their interaction. The analytic process by which both of these methodologies expose such conversational regularities is centered around the identification and analysis of communicative breakdown, making them strong foundations for evaluating the communicative efficacy of interaction.

2.3.1 Conversation Analysis: Exposing Conversational Regularities

The central goal of conversation analytic studies is to expose and document the competences, or *conversational regularities*, that interacting participants rely on to make available their interpretations of ongoing action and organize the communicative behaviors. The basic premise (Garfinkel, 1967) is that ordinary talk is a highly organized phenomenon and that all communicative activities, from producing communicative displays to interpreting those of others, can be accounted for as products of a common set of such conversational regularities. That is, the words that appear in an utterance are not viewed as semantically rich symbols, but as linguistic resources for negotiating the illocutionary significance of the interaction as a whole (e.g. accusations, complaints, requests, etc.). The way in which the utterance is designed to achieve such illocutionary goals is informed by the underlying body of conversational regularities, which capture the organized procedures, interpretative expectations, and resources which are known to speakers by virtue of long-standing membership in a language-using community.

It is important to emphasize that conversational regularities do not embody deterministic heuristics for interpreting the significance of action by universally defining the precise meaning of specific communicative displays. Rather, they establish an overall framework of socially-established conversational practices that participants rely on to inform the local, contextual construction of the significance of each other's communicative behaviors. For example, it has been observed (Fox, 1993; Pomerantz, 1975; Pomerantz, 1978) that participants regularly use extended pauses to avoid explicitly producing a critical or negative response to a partner's immediately preceding utterance. In orienting to this conversational regularity, the speaker who produced the original utterance is able to recognize that a socially dispreferred response may be imminent, and is provided with an opportunity to retract or rephrase the utterance. The fact that conversational participants regularly exhibit this behavior, however, does not mean that *every* extended pause necessarily signifies a negative response. There may be any number of unique local contingencies (e.g. one's partner is attending to something else) that result in an alternative interpretation of an extended pause. In this way, conversational regularities abstractly characterize patterns of behavior consistently used by conversational participants to achieve their communicative goals, but do not deterministically describe the significance of action. In particular, whether or not a given conversational regularity applies in a specific situation (e.g. whether a *particular* extended pause implies an imminent socially dispreferred response) can only be determined by participants *in situ*, with respect to the unique local context of the evolving interaction.

This fundamental distinction between heuristic interpretation and regularities in communicative behavior is stressed by Wooffitt:

> [Conversational regularities], then, are normative, socially organised procedures. They are not 'hard wired' into cognitive processes in such a way that they determine or propel the turns that people produce in interaction. Nor do they exist independently of those occasions in which their relevance is oriented to by participants in conversation. Rather, they are instantiated in the local, turn-by-turn particulars of interaction. They are contingent upon, and realised through, people's orientation to their normative or programmatic character. (Wooffitt, 1990, p. 27)

This commitment to the contextual construction of significance places the analytic focus of Conversation Analysis squarely on the empirically-observable communicative behaviors of participants. Instead of speculating about what the participants might "know" and how this could conceivably influence interaction, the analysis works to document the regularities that participants consistently orient to over the course of naturally-occurring interactions.

The pragmatic analytic processes that comprise Conversation Analysis are summarized in Figure 2.6.

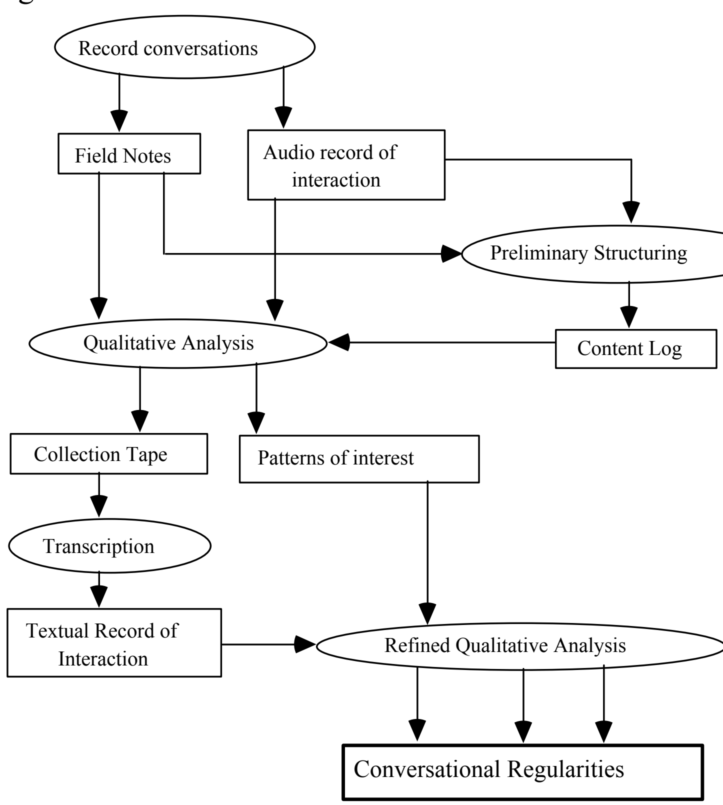


Figure 2.6: Overview of Conversation Analysis. Ovals represent analytic processes, rectangles represent analytic results.

As indicated in Figure 2.6, the analysis is empirically grounded in audiotape recordings of naturally occurring interactions. Unlike traditional scientific methods, which tend to rely on statistical sampling theory, the selection of interaction to examine by Conversation Analysis is usually thematic, shaped by the analyst's interests. For instance, researchers may focus on certain types of interactions, on certain people or classes of people, or on particular objects or documents around which a task is organized (Suchman & Trigg, 1991) . This focus on naturally occurring interactions and on thematic sampling shaped by the social realities of mundane communicative interaction reflects an underlying commitment to the ideals of Situated Action, namely, that meaning and proof reside only in the situated circumstances of action.

In subsequent steps of the analysis, the audio record is intensively reviewed to develop an increasingly refined understanding of how specific communicative behaviors are used by participants to display their interpretations of action and how those displays are perceived by conversational partners. After developing a content log that establishes a rough temporal map of the interaction and summarizes its structure, the audio records are analyzed in detail, yielding a preliminary articulation of the conversational regularities that participants use to organize their interaction. Analysis is focused by a well-defined set of issues or *analytic foci* (Jordan & Henderson, 1995) : How do events begin and end? How is activity organized temporally? What are the mechanisms of breakdown and repair?

In order to reduce the chance of individual bias, this stage of the analysis is often a collaborative effort between several researchers, ensuring that multiple interpretations are explored. Interesting patterns of behavior emerging from this analysis are then further examined by creating a collection tape which brings together sequences that instantiate those patterns, and then carefully transcribing the tape to create an easily-examinable textual record of interaction. These transcripts are then subjected to further analysis and eventually serve as empirical evidence for the conversational regularities identified as the end result of the analysis.

Conversation Analysis has been used to gain insight into a wide variety of conversational regularities that participants use to organize various aspects of their interaction; examples include studies of how participants regulate contributions to the interaction (Jefferson, 1988; Jefferson, 1992; Sacks, Schegloff, & Jefferson, 1974) , maintain a shared topical orientation (Beach, 1993; Button & Casey, 1984; West & Garcia, 1988) , and manage repair (Fox, 1993; Schegloff, Jefferson, & Sacks, 1977) .

2.3.2 Interaction Analysis: Attending to Nonverbal Behaviors

Historically, conversation analytic studies have concentrated mainly on the verbal aspects of interaction, often focusing on telephonic conversations, in which participants' nonverbal displays are irrelevant. Even when studies have examined conventional face-to-face interactions, nonverbal displays have generally been accorded much less detailed attention than the utterances produced by participants (Schegloff, 1972; Schegloff, 1979; Schegloff & Sacks, 1973) . This observation does not imply that non-verbal displays are not thought by conversation analysts to play an important role in organizing interaction. Rather, the assumption is that the verbal channel is the *primary* means of organizing conversational activity. More pragmatically, it has been suggested that the focus on verbal activity in much early work in Conversation Analysis represents a way of taming the overwhelming complexity of communicative interaction:

> ...it does not follow that conversation analysts are therefore uninterested in or content to ignore the possible significance of non-vocal activities. Indeed, the widespread use that has been made of recorded telephone calls as a focus for analysis recognizes a major methodological advantage precisely in the fact that the participants themselves cannot see each other. The analyst can thus proceed to the study of audio recordings without having to worry about how non-vocal activities may have been involved in a particular sequence. (Atkinson & Heritage, 1984, p. 223)

While the focus on verbal content may be justified in copresent or telephonic interactions (but see Whalen, 1995) , consideration of nonverbal behavior is vitally important for evaluating the efficacy of interaction in technologically-mediated environments. In the analysis of copresent interaction, it is possible to argue that, *all other things being equal*, verbal communication is the participants' primary resource for organizing their communicative behavior. That is, nonverbal activity can be viewed as a sort of "constant" in the conversational equation. But when comparing interactions in different technologically-mediated environments, all things are *not* equal because the simulation of copresence defined by each electronic environment uniquely compromises the communicative resources available in copresent interaction; some resources are affected more than others. For instance, the fixed frame and resolution limitations of the stationary cameras used to create a video connection in many technologically-mediated environments (e.g. Dykstra-Erickson, Rudman et al., 1995) clearly constrains the visual information available to participants in different ways than environments (Gaver, Smets, & Overbeeke, 1995) that support multiple views. In general, different technologically-mediated environments may provide widely different levels of both audio and video quality; expanding the traditional conversational analytic techniques to take into account non-verbal behaviors is crucial for understanding how the communicative efficacy of participants' interactions is affected by such variations.

These observations, coupled with the increasing availability of sophisticated video equipment suitable for analyzing videotaped interactions, have served to motivate an increasing trend towards including participants' nonverbal displays in more recent conversation analytic studies. Several analytic efforts (Goodwin, 1981; Goodwin, 1986; Heath, 1986; Schegloff, 1984; Suchman, 1987) have extended the notational conventions traditionally used by conversation analysts to denote certain non-verbal behaviors, and to take these behaviors into account in the subsequent analysis. This gradual evolution from the analysis of audio-only interactions to audio-video interactions, and the accompanying expansion of analytic focus to consider nonverbal as well as verbal behaviors has recently been formally recognized by coining the term *Interaction Analysis* to refer to conversation analytic studies of videotaped interactions (Jordan & Henderson, 1995) . The underlying goals and analytic processes of Interaction Analysis are identical to those of Conversation Analysis (see Figure 2.6); the only difference is that the conversational regularities revealed by the analysis may also include regularities in the way participants utilize nonverbal displays as resources for organizing their interaction.

In sum, Conversation and Interaction Analysis essentially embody the same analytic process, centered around the empirical analysis of naturally occurring communicative interactions to expose conversational regularities that interacting participants rely on to organize their communicative displays and shape interpretation of the displays of their conversational partners. While Conversation Analysis has historically tended to focus primarily on the verbal components of interaction, the recent development of Interaction Analysis extends conversation analytic techniques to take into account non-verbal behaviors as well. Because both techniques are based on the retrospective reconstruction of the significance of events, Conversation and Interaction Analysis inevitably requires some amount of inference on the part of the analyst. However, a strong emphasis is placed on grounding all such subjective assessments in raw evidence.

2.3.3 Interaction Analysis as a Tool for Exposing Breakdown

In the preceding discussion, Interaction Analysis was introduced as a methodology for documenting the underlying conversational regularities that conversants rely on to shape the production and interpretation of each other's verbal and non-verbal communicative displays. However, the analytic procedure used by Interaction Analysis to reveal these regularities distinguishes Interaction Analysis as a strong methodological foundation for evaluating the efficacy of interaction as well. Specifically, the approach used by interaction analysts to expose conversational regularities is based on the idea that such regularities only become apparent to the analyst when they are somehow violated, resulting in communicative breakdown. In other words, the only way to show that a regularity does, indeed, exist and is oriented to by interacting participants, is to highlight the communicative troubles that result when the regularity in question goes unrecognized. This essential role of communicative breakdown in the analytic process is emphasized by Atkinson and Heritage:

> Generally, the analyst will also take steps to demonstrate that the regularities are methodically produced and oriented to by participants as normatively oriented-to grounds for inference and action. As part of this latter objective, the analysis of "deviant cases" — in which some proposed regular conversational procedure or form is *not* implemented or realized — is regularly undertaken. (Atkinson & Heritage, 1984, p. 2, original emphasis)

This relationship between communicative breakdown and underlying conversational regularities is depicted in Figure 2.7.

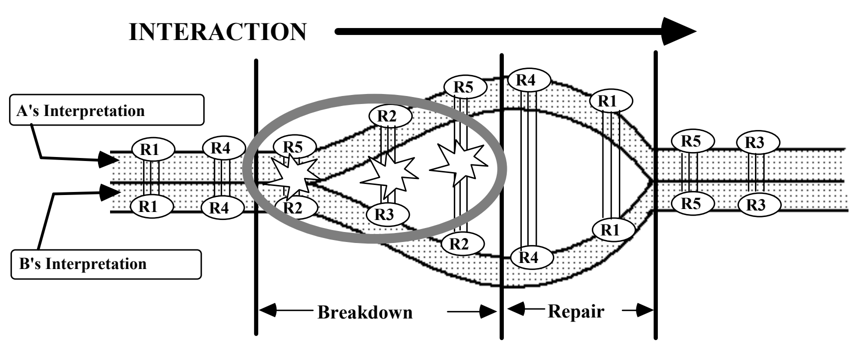


Figure 2.7: Communicative breakdown draws attention to and confirms the existence of underlying regularities by concretely evidencing the communicative consequences of violating them. The ovals labeled R1, R2, etc. represent various regularities oriented to by participants.

Figure 2.7 reifies the earlier discussion of conversational regularities, illustrating how participants continuously rely on such regularities to inform the production of communicative displays, and the interpretation of the displays of others, thereby maintaining shared interpretations of the emerging significance of the interaction. Communicative breakdown results from a failure to mutually recognize a given regularity, resulting in divergent interpretations of behavior; confusion continues until the breakdown is detected and somehow repaired.

To illustrate these ideas, consider the following example. It has been observed (Sacks, 1992b; Sacks, Schegloff et al., 1974) that conversational participants rely on a relatively simple set of conversational regularities to organize their verbal contributions to an interaction, passing control of the conversational floor back and forth between them to avoid overlapping talk. One such regularity is that a speaker's turn at talk can generally be assumed to have ended after the speaker asks a question, providing a conversational partner with an opportunity to respond. There are two ways that this conversational regularity becomes apparent to the interaction analyst: (1) through the empirical observation that control over the conversational floor regularly passes to a new speaker after a question is asked, and (2) by documenting cases of communicative breakdown that occur when either the speaker or the listener fail to orient to this regularity for some reason, resulting in a divergent interpretation of whose turn it is to speak. If the speaker continues on, his or her utterance may overlap with the response produced by the listener; if the listener fails to recognize the turn ending, the resulting silence may, in itself, be misinterpreted as a response (e.g. as the avoidance of a socially dispreferred response), causing further confusion.

In this way, the analysis of communicative breakdown is a vital technique for exposing the underlying conversational regularities that participants use to inform their interaction. By documenting the communicative troubles that result when conversational regularities are somehow violated, the analysis demonstrates that such regularities actually exist and are important as mechanisms for organizing interaction.

Clearly, the practical focus in Interaction Analysis on exposing communicative breakdown provides a strong methodological foundation for evaluating the communicative efficacy of interaction. As we shall see in the following chapter, however, the essentially documentary goals of Interaction Analysis prevent us from using it directly to compare the overall communicative efficacy of interactions, motivating substantial modification and extension of the methodology to create an appropriate analytic tool for accomplishing the comparative analysis presented in this work.

2.4 Summary: Theoretical Foundations

The goal of this chapter has been to establish the theoretical underpinnings for the analytic technique used to compare the communicative efficacy of copresent and technologically-mediated interaction in this dissertation. Situated Action was introduced to establish the epistemological foundations for this research, providing a solid basis for understanding what it means for communicative interaction to be effective, and how the efficacy of interactions might be evaluated. Under Situated Action, communication was characterized as a collaborative negotiation of significance, in which participants continuously monitor and interpret the communicative displays of others, while at the same time making available evidence of their own evolving interpretation of mutually available events. Accordingly, the communicative efficacy of interaction is defined by the extent to which participants are able to maintain intersubjectivity throughout their interaction, avoiding divergent interpretations of action, or *communicative breakdowns*.

Interaction analysis was introduced as a strong methodological foundation for revealing communicative breakdown. Though the primary goal of Interaction Analysis is to document certain conversational regularities that participants rely on to organize their communicative activities, a key technique used to reveal these regularities is centered around exposing and analyzing communicative breakdowns experienced by participants. This focus on exposing communicative breakdown makes Interaction Analysis a promising foundation for the comparative analysis of communicative efficacy undertaken in this project.

The main points of this chapter can be summarized as follows:

1. Communication is properly characterized as a situated, evidentiary process in which interacting participants continuously work to construct shared interpretations of ongoing action.

2. Communicative trouble is characterized as a failure of participants to maintain intersubjectivity, resulting in divergent interpretations of interaction evidenced by observable confusion and repair.

3. The communicative efficacy of an interaction is defined by the extent to which participants are able to maintain shared interpretations of action. Pragmatically, it can be defined by the amount of communicative breakdown experienced by participants over the course of their interaction.

4. Interaction analysis provides a promising methodological foundation for evaluating the communicative efficacy of interaction, in that it is based on exposing patterns of communicative breakdown.

In the following chapter, we apply the ideas presented here to develop a methodology for comparatively evaluating the communicative efficacy of interactions taking place in different communication environments, and formally introduce the comparative analysis of copresent and technologically-mediated interaction that is the focus of this work.

2.5 Notes

[1] The term *intelligent agents* is purposefully broad, since Representationalism draws no distinction between humans and other symbolic processors.

[2] There is no value to a symbolic representation whose "truth value" is uncertain. That is, there's no point in representing anything if the meaning of the representation cannot be unambiguously established. This observation is reflected in the obsession with formal semantics in the knowledge representation community (Brachman & Levesque, 1985) .

[3] Vera and Simon (1993) have suggested that the dynamic, situated way in which rational action arises under Situated Action can be seen as merely a particular fine-grained instantiation of the model-based reasoning processes underlying the Representationalist view.

CHAPTER III

METHOD

The goal of this chapter is to set the cornerstone for the comparative analysis of copresent and technologically-mediated interaction presented in this work by formally describing the methods used in the analysis. Drawing on the methodological foundations established in Chapter II, we begin by introducing the methodology used to evaluate and compare the communicative efficacy of interactions in various communication environments. We then turn to a thorough description of the interactions that were arranged to generate the data for the comparative analysis, including the three communication environments that were compared, the participants, the tasks they were presented with, and the arrangements for data collection. The closing sections discuss the analytic steps taken to apply the methodology described earlier to the data, setting the stage for the presentation of results in the following chapters.

3.1 Analytic Tools: Why Not Use Interaction Analysis?

In Chapter II, Interaction Analysis was identified as a methodological foundation for understanding the communicative efficacy of interaction, in that it is designed specifically to reveal the communicative difficulties encountered by participants. In particular, Interaction Analysis directly examines the communicative interaction of participants to reconstruct the collaborative process by which participants are able to construct shared interpretations of action; the communicative breakdowns exposed by this analysis can be viewed as a direct measure of the communicative efficacy of interaction.

Despite these promising foundations, Interaction Analysis does not constitute a suitable analytic tool for the comparative analysis of communicative efficacy because its analytic orientation is fundamentally *documentary* rather than comparative. That is, interaction analysts are not fundamentally interested in communicative breakdown and the efficacy of interaction, but rather in the communicative regularities that participants use to regulate their interaction and make available their interpretations of mutually available events. Exposing communicative breakdowns that occur during an interaction is not the primary goal of the analysis, but merely represents a means to this end — one way in which communicative regularities are revealed to the analyst is when they are violated in some way and result in breakdown. In short, breakdowns are interesting to interaction analysts because they are useful as a means of exposing and documenting the communicative regularities in interaction rather than as a metric of communicative efficacy.

The principle consequence of this difference in analytic emphasis is that, while it does expose the *kinds* of breakdown that occur, Interaction Analysis provides no basis for assessing the total *amount* of breakdown that exists in an interaction. As a result, it provides no basis for quantitatively analyzing differences in the amount of breakdown that exist between interactions occurring in different communication environments in order to assess their relative communicative efficacy.

From the practical standpoint of designers working to create more robust simulations of copresence, the inability of Interaction Analysis to reveal differences in the prevalence of breakdown between environments also compromises its power to inform future designs. To see this, suppose that Interaction Analysis were applied to interactions that took place in two different environments, revealing several interesting patterns of communicative breakdown that occurred. Which of these patterns of breakdown are related to differences in the overall communicative performance of the two environments and are therefore deserving of further analysis? Without a quantitative characterization of which patterns of breakdown were significantly more prevalent in one environment than in the other, there is no principled way to answer this question.

In sum, the fact that Interaction Analysis is, at heart, a *documentary* rather than a *comparative* methodology makes it unsuitable for the comparative analysis of communicative efficacy undertaken in this work. In particular, the fact that Interaction Analysis does not support the quantitative generalization and comparison of communicative efficacy between different environments means that it can not expose the relative communicative efficacy of technologically-mediated and copresent interaction, or provide any insights as to why differences in communicative efficacy exist.

3.2 Analytic Tools: Introducing Breakdown Analysis

The methodology developed for this project extends Interaction Analysis in several ways to create a focused analytic tool for evaluating and rationalizing the communicative efficacy of communication environments. We refer to this methodology as *Breakdown Analysis*, reflecting the emphasis on communicative breakdown as a metric for assessing communicative efficacy. Breakdown Analysis extends Interaction Analysis in three important ways:

1. Completeness. Breakdown Analysis is based on the premise that the communicative efficacy of interactions is directly reflected in the total amount of communicative breakdown experienced by participants. Consequently, it is necessary to analyze *entire* interactions, exposing all of the breakdowns that occur, rather than just focusing on segments that reveal the existence of certain patterns of breakdown.

<u>2. Quantitative Comparison.</u> Breakdown Analysis supports the statistical analysis of differences in the amount of breakdown between interactions occurring in different communication environments by revealing both patterns of breakdown and the total amount of breakdown in each of these categories.

<u>3. Rationalizing Differences.</u> Breakdown Analysis supports the focused investigation of why differences in communicative efficacy exist, by providing both a concise articulation of communicative troubles encountered by participants, and differences in the prevalence of such troubles in different communicative environments.

The differences between Interaction Analysis and Breakdown Analysis are illustrated in Figure 3.1, which provides a graphical comparison of the two techniques.

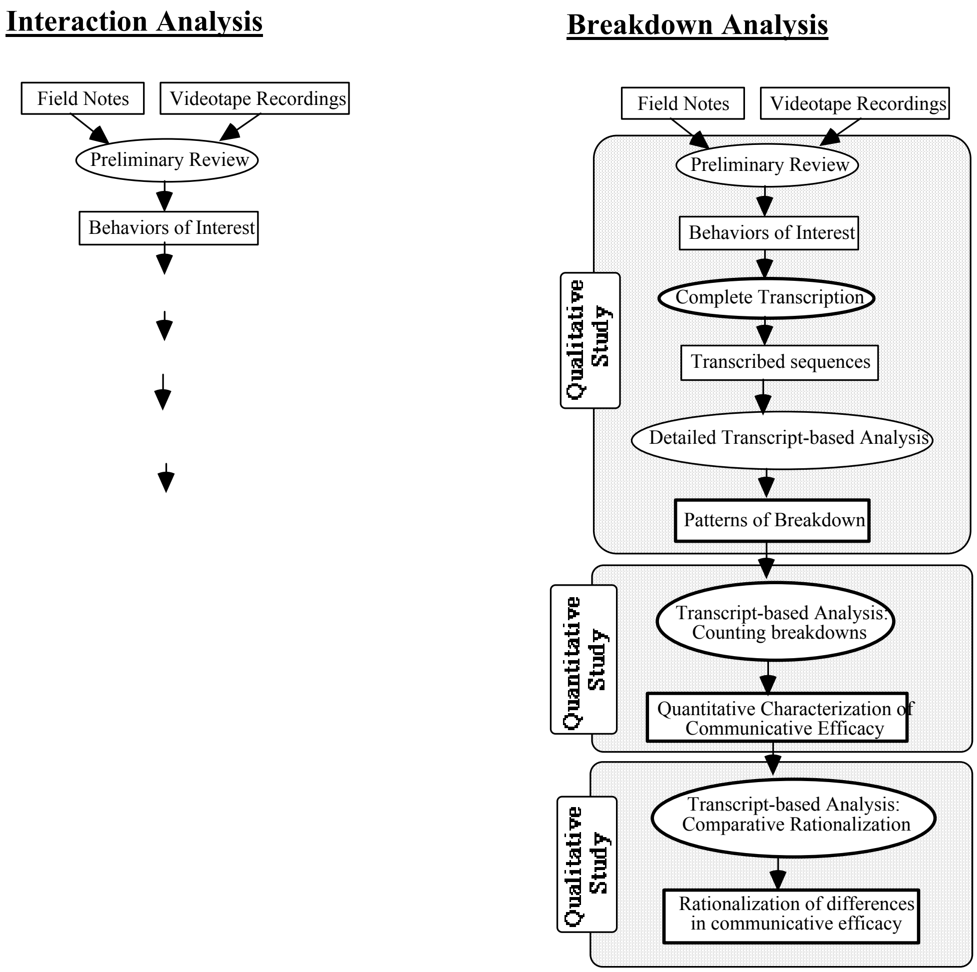


Figure 3.1: Comparison of (a) Interaction Analysis to (b) Breakdown Analysis. Ovals represent analytic processes; boxes represent analytic products.

As indicated in Figure 3.1, the initial stages of Breakdown Analysis and Interaction Analysis are essentially identical, focused on exposing interesting patterns of behavior through a process centered around the review and eventual transcription of videotaped interactions. These transcripts then serve as a foundation for detailed analysis of the patterns exposed in the preceding analysis. However, the goals of this initial qualitative analysis differ for the two techniques. Where Interaction Analysis is aimed at exposing individual episodes of breakdown as a way of revealing underlying communicative regularities, the goal of Breakdown Analysis is to expose consistent patterns of communicative breakdown. The eventual aim in Breakdown Analysis to quantify the total amount of breakdown in each interaction gives rise to a more practical difference: each interaction must be transcribed in its entirety.

As shown in Figure 3.1, Breakdown Analysis then extends Interaction Analysis with two further studies aimed, respectively, at stochastically comparing the frequency of breakdown between interactions occurring in the different communication environments being compared, and rationalizing the differences in communicative efficacy exposed.

In this way, Breakdown Analysis can be characterized as a data collection effort followed by three interleaved qualitative and quantitative studies, which progressively refine our understanding of the communicative troubles experienced by participants in each of the communication environments being compared. The following paragraphs present a functional overview of each stage of the analysis; a discussion of the practical aspects of each stage is reserved for section 3.4, where we discuss analytic procedures in more detail.

3.2.1 Qualitative Study #1: Identifying Patterns of Breakdown

The goal of the first qualitative phase of Breakdown Analysis is to develop the comparative framework for the entire analysis by identifying consistent patterns of communicative breakdown. Participants are videotaped[1] interacting in each of the communication environments being compared, and the videotapes then subjected to intensive review to gain a preliminary understanding of the types of communicative difficulties that participants experienced over the course of interactions. These observations are used as a basis for developing a notational schema for transcribing the interactions which faithfully documents the verbal and non-verbal behaviors of participants, while avoiding unnecessary detail that might obscure the analysis. All of the recorded interactions are then transcribed in their entirety. Finally, these transcripts are then used as the basis of further in-depth analysis, aimed at refining the categories of breakdown identified during the earlier review of the videotape data, developing strong and consistent evidentiary criteria for recognizing breakdowns in each category. The categories of breakdown identified through this analysis establish the dimensions on which the communicative efficacy of interactions will be compared.

3.2.2 Quantitative Analysis: Exposing Differences in Efficacy

The goal of this phase of the analysis is to quantitatively compare the amount of breakdown documented in different communication environments. Using the operational criteria defined in the preceding stage of the analysis, each interaction is analyzed in its entirety to expose all of the breakdowns in each category that occurred. This characterization of the total amount of communicative trouble that occurred in each interaction is used as the basis for comparing the overall communicative efficacy of interactions taking place in different communication environments. A statistical comparison of the number of breakdowns in each category is performed to reveal significant differences in the frequency of breakdown between environments which, in turn, are taken to directly reflect differences in communicative efficacy that exist between environments.

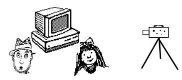
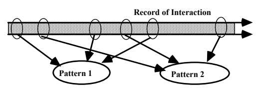
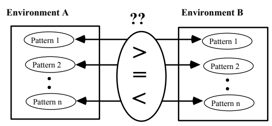
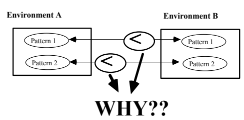
3.2.3 Qualitative Study #2: Rationalizing Differences in efficacy

The goal of this final phase of the analysis is to rationalize observed differences in communicative efficacy by establishing causal relationships between certain physical characteristics of technologically-mediated environment and the communicative breakdowns that occurred in those environments. Because the commitment to the underlying epistemology of Situated Action rules out context-independent explanations that deterministically link communicative failure to specific features of the communication environment, the analysis is aimed at revealing the ways in which the design of an environment constrains participants' access to vital communicative resources (e.g. eye gaze, gestures, hand position, audio cues, and so on), thereby compromising participants' ability to monitor each other's evolving interpretations of the interaction and increasing the *likelihood* of breakdown.

3.2.4 Summary: Breakdown Analysis

Breakdown Analysis modifies and extends the well-established methodology of Interaction Analysis to create an analytic tool for comparatively evaluating the communicative efficacy of interaction in different communication environments. Breakdown Analysis relies on the qualitative techniques of Interaction Analysis to identify categories of communicative breakdown that exist in interactions, then enumerates the total number of breakdowns that occurred in each category. A statistical comparison of breakdowns between interactions that took place in different environments is used to determine the relative communicative efficacy of those environments. Finally, the concise articulations of communicative troubles and the differences in the prevalence of such troubles yielded by the analysis are used to focus a further qualitative study aimed at explaining why observed differences in efficacy exist, by revealing impediments to participants' successful interpretation of each other's communicative displays imposed by the environment. The methodology of Breakdown Analysis is summarized in Table 3.1:

Table 3.1: Overview of Breakdown Analysis.

Phases of Analysis		Method	Purpose/ Results
Data Collection		Videotape interactions in each environment	Create data for analysis
Qualitative Study #1		Interaction Analysis: Review, transcribe, expose breakdowns	Expose consistent patterns of breakdown
Quantitative Study		Statistical Analysis: Statistical comparison of amount of breakdown	Reveal differences in communicative efficacy
Qualitative Study #2		Interaction Analysis: Focused examination of breakdowns.	Explain differences in communicative efficacy

As indicated in Table 3.1, Breakdown Analysis intertwines the qualitative techniques of Interaction Analysis with the quantitative techniques of traditional scientific methods to yield insights into the functional differences that exist between communication environments, and how those differences are related to the design of the environment. As such, it provides the ideal tool for performing the comparative analysis of copresent and technologically-mediated interaction undertaken in this work.

More generally, Breakdown Analysis represents an application of Exploratory Sequential Data Analysis (ESDA), a term which has been recently introduced (Sanderson & Fisher, 1994) to characterize a broad range of analytic techniques[2] developed to expose meaningful patterns in sequentially organized data. Unlike tradition scientific approaches, which are based on the notion of hypothesis testing and confirmation, ESDA techniques focus more on hypothesis formation.

3.3 Comparing Copresent to Technologically-Mediated Interaction

The goal of this dissertation is to explore the extent to which technologically-mediated communication environments support the same communicative efficacy as copresent interaction. So far, the focus has been on establishing the prerequisites for this comparative analysis, by defining a theoretical basis for understanding the notion of "successful" communication, and developing analytic tools for exposing and comparing the communicative efficacy supported by different communication environments. With the introduction of Breakdown Analysis in the previous section, we are finally equipped to conduct a comparative analysis of copresent and technologically-mediated interaction. This section describes how the interactions used in the analysis were generated and recorded. We begin with a brief overview of the study; following sections are devoted to detailed description of the environments, tasks, participants and arrangements for data collection.

3.3.1 Overview of the Study

Pairs of participants were videotaped while collaborating to accomplish a simple analytic task in three different communication environments: face-to-face, distributed while connected by audio link, and distributed while communicating by audio/video link. The task involved using a simple but powerful cardiovascular simulator to construct a dynamic representation of a given problem statement and use that representation to answer a series of questions. Participants were naive computer users with no previous teleconferencing experience, and were unfamiliar with the simulator as well. In all three communication environments, participants shared access to an electronic workspace containing the running simulator and were both able to manipulate the simulator by way of a shared cursor in the workspace. Four pairs of subjects were recorded in each of the three communication environments, yielding a total of twelve interactions.

3.3.2 Justifying the Experimental Design

In any scientific effort, the validity of results is critically dependent on the design of the experiment used to generate them. In the context of the comparative analysis presented here, the central issue is how to choose the specific communication environments to compare in such a way as to maximize the generality and applicability of the results. Each facet of this experiment — the environments, the participants, and the task — were carefully chosen to accentuate differences in communicative efficacy that exist in a set of environments very similar to ones that are currently becoming available to the general public. Briefly, the design of the experiment is rationalized as follows:

Environments. The primary reason for choosing audio-only and audio-video environments for comparison against copresent interaction is that they represent the basic choices of media available to designers of current technologically-mediated environments. Consequently, insights or limitations regarding the technologies and techniques used to implement these environments will be of interest to a broad audience. Second, comparison of the audio-only and audio-video conditions directly tests the Bandwidth Assumption underlying the design of current technologically-mediated environments since, clearly, the audio-only environment supports a much lower bandwidth than the audio-video environment. If the Bandwidth Assumption holds, the comparative analysis presented here should show that interactions in the audio-video condition have a significantly higher communicative efficacy than those in the audio-only condition.

Task and Participants. Both the task and the participants used in this analysis were chosen specifically to place extraordinarily high demands on the communication environment by greatly increasing the level of collaborative interaction between participants, thereby accentuating differences in communicative efficacy that exist between environments. For instance, the fact that the participants are computer-naive and have no experience with the simulator guarantees substantial confusion, resulting in a great deal of discussion as participants work to accomplish the task. In learner-learner interactions, both participants have only weak conceptions of basic domain ontology, no clear notion of what constitutes a solution to a given instruction, and no shared foundations in the customs and techniques associated with an expert community of practice. The choice of an *interpretive* rather than a constructive task further exacerbates matters. When participants are engaged in a constructive task like, for example, shared drawing, at least one participant (i.e. the one who draws them) understands the significance of the symbolic representations being produced in the shared workspace. By contrast, neither of the participants interpreting the behavior of a dynamic simulation has any *a priori* understanding of the significance of the symbolic presentations of the simulator.

In sum, the analysis presented in this work is based on a comparison of the communicative efficacy of interactions between inexperienced participants performing a novel analytic task in three very different communication environments. The following sections describe each facet of the experiment in more detail.

3.3.3 Environments: Three Communicative Conditions

The three environments in which interactions took place represent the independent variable in the comparative analysis presented in this study. A descriptive overview of the three communicative conditions is presented here; Appendix B provides an in-depth discussion of the technical arrangements used to implement each environment.

Copresent Environment. In the copresent environment, participants were seated side-by-side in front of a single screen displaying the simulator workspace. Participants were able to speak, gesture and point freely; they had completely natural access to each other's verbal and non-verbal behaviors. The copresent condition provides the analytic baseline for the comparative analysis presented in this study, setting the standard for communicative efficacy against which interactions in the two technologically mediated environments are evaluated.

Audio-Only Environment. Participants interacting in the audio-only environment were placed in front of individual monitors in separate rooms. Shared access to the simulator workspace was provided by splitting the screen output of a single workstation, amplifying the signal, and sending it to the two monitors viewed by participants. This reliance on straightforward analog technology[3] made it possible to provide high-quality, latency-free access to a shared electronic workspace. A high-fidelity audio connection between participants was provided by equipping each participant with a lapel microphone and a lightweight headset. The audio circuit was designed to mix the inputs from the two participants and then distribute the resulting signal, yielding a connection functionally similar to a telephone connection, though of substantially higher audio quality.

Audio-Video Environment. The audio-video environment was identical to the audio-only environment, except that a video connection between participants was added. Each participant was provided with a large video monitor displaying an image of the remote participant[4]. The monitors were placed adjacent to the screen displaying the simulator workspace, separated by about 50 degrees of angular displacement. The camera recording the participants' behavior was placed within this angle, elevated to provide the best possible view of the remote participant. The specifics of camera placement are presented in section 3.3.6, where we discuss arrangements for data collection in more detail.

3.3.3.1 Acting in the shared electronic workspace

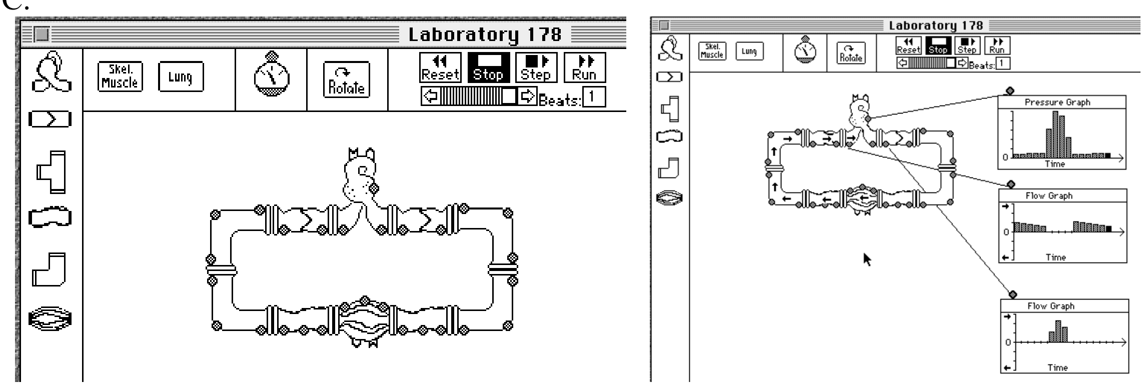
In addition to having visual access to the electronic workspace, participants were also empowered to act in the workspace by providing them each with their own mouse. However, there was only a single cursor available within the shared workspace, which was continuously controlled by both mice. This feature resulted in erratic[5] cursor behavior when both participants attempted to simultaneously move the cursor. Since no formal mechanisms for regulating access to the shared cursor were provided by the system, it was incumbent on participants to organize their contributions using the cursor in such a way as to avoid these disruptive "cursor wars."

3.3.4 Task

The task performed by participants in each interaction involved the manipulation and interpretation of a simple cardiovascular simulator called the Cardiovascular Construction Kit[6], or CVCK (Douglas & Doerry, 1994b) . This section provides a brief overview of task structure and content; Appendix A gives a detailed description of the CVCK.

The CVCK simulator allows users to explore the behavior of arbitrary circulatory constructs by piecing together primitive cardiovascular components defined by the CVCK (e.g. ventricles, valves, muscles, lungs, and capillary beds) and then running the simulation to reveal the behavior of the cardiovascular system constructed. Gauges may be attached to record and display the behavior of certain parametric quantities (e.g. pressure, flow, etc.) at various points in the construction.

Participants were asked to use the CVCK simulator to collaboratively work through a series of exercises described in a "laboratory manual." To complete the task, participants had to first construct a simple cardiovascular loop depicted in the laboratory manual, and then run the simulation to analyze its hydraulic behavior. They were then asked to attach gauges to reveal further behavioral detail and, again, use the simulator to answer a second series of questions. The sequence of constructions produced is shown in Figure 3.2; the complete laboratory manual given to participants is available in Appendix C.



(a) (b)

Figure 3.2: Construction Tasks: (a) Build a basic valved loop and (b) Add gauges to measure values

The interleaved construction and analysis steps in this process provide a basis for distinguishing four distinct phases in the task-solution process. Table 3.2 summarizes these four phases, and highlights the main issues that participants had to resolve in each phase.

Table 3.2: Overview of task solution steps and issues to be resolved by participants.

Step in task solution	Issues to resolve
Step 1: Initial Construction Construct simple valved loop starting with blank workspace.	• How to create new components. • How to rotate the “elbow” component used for the corner pieces in the cardiovascular loop.
Step 2: Interpretation Run the simulator, interpret its dynamic behavior to answer questions posed in laboratory manual.	• Which components are the “valves”?
Step 3: Attaching Gauges Attach gauges to the points on the construction specified in the laboratory manual.	• What are “gauges?” • How to attach gauges. • How to differentiate between pressure and flow gauges.
Step 4: Interpretation Run the simulator again to answer further questions regarding its dynamic performance.	• Which component is the “heart?” • How to interpret gauge presentations.

The list of issues shown in Table 3.2 emphasizes an important feature of the laboratory manual given to participants: It was extremely minimalist in nature. Rather than walking learners through the exercises step by specific step, the manual given to participants specified only broad objectives. For example, participants were instructed to “Attach gauges at the places marked in [the figure shown in your laboratory manual],” but were not told which of the icons in the simulator workspace represents a gauge, or how one might go about attaching one. This minimalist approach ensured that participants would face plenty of quandary and confusion, promoting extensive discussion and close collaboration as participants worked to figure things out.

Participants were asked to mark their answers to questions directly in the laboratory manual. In the copresent condition, the two participants shared a single laboratory manual since they were seated adjacent to one another; in the two distributed conditions, each participant was given a copy of the laboratory manual.

3.3.5 Participants

Participants were recruited from three undergraduate biology classes[7] at the University of Oregon: General Biology, Human Physiology, and Anatomy. This choice of subject population reflects a concerted effort to produce more naturally motivated interactions by recruiting participants who have strong personal reasons to be interested[8] in the task domain (i.e. cardiovascular physiology). Participants were encouraged to sign up in pairs, by selecting a friend as a partner. This self-paired approach to recruitment has strong practical motivations as well: Past experience with protocol analysis (Douglas & Doerry, In preparation) has shown that communication between self-paired participants is much less inhibited than between strangers. Specifically, we have noted that differences in social status are less apparent among self-paired groups, and that they feel more comfortable critiquing each other's decisions.

Potential participants were given a short questionnaire (see Appendix C) to fill out to collect contact information and to establish their educational history and computer-related background. Based on this information, pairs of participants were selected according to the following priorities:

1. No pairs in which either participant had any substantial experience with interactive telecommunications outside of mundane person-to-person telephony were allowed, in order to exclude participants who might have established pre-existing competency in technologically-mediated interaction.

2. Pairs in which both participants had similar amounts of experience using computers were given priority. The idea was to avoid a situation in which one much more experienced participant dominated the entire interaction.

Twelve pairs of participants were recruited to fill the needs of the analysis. No effort was made to control the gender balance of the participants recruited; seven female-female pairs, three male-male pairs, and two male-female pairs were eventually selected. Subjects were paid $5 for their participation.

3.3.6 Creating Records of Interaction

Because a complete, unbiased and unobtrusively produced record of the interaction is the foundation for Interaction Analysis and, by extension, Breakdown Analysis, the importance of making careful arrangements for videotaping participants' interactions can not be overstated. The value of a strong audio and video record is heavily emphasized in Jordan and Henderson's treatise on Interaction Analysis (Jordan & Henderson, 1995) and conclusively illustrated by a number of existing analyses (Fox, 1993; Heath, 1986; Heath & Luff, 1992; Heath & Luff, 1993; Suchman, 1987; Suchman & Trigg, 1991; Tang, 1991) .

Unlike an audio record, which is more or less omnidirectional, the content of a video image is absolutely determined by the placement of the camera and how the image is framed. In general, discussion of these issues is shaped by an inherent tradeoff between field of view and the amount of detail available in the recorded image; a wider field of view captures a larger physical space, but inevitably compromises the ability to discern fine-grained details in the image. For example, framing the image to capture both the participants and their immediate surroundings implies that details of facial expression may not be discernible in the video record. Consequently, the analyst must carefully consider what visual aspects of the interaction are likely to be most important to the subsequent analysis before deciding on appropriate camera arrangements.

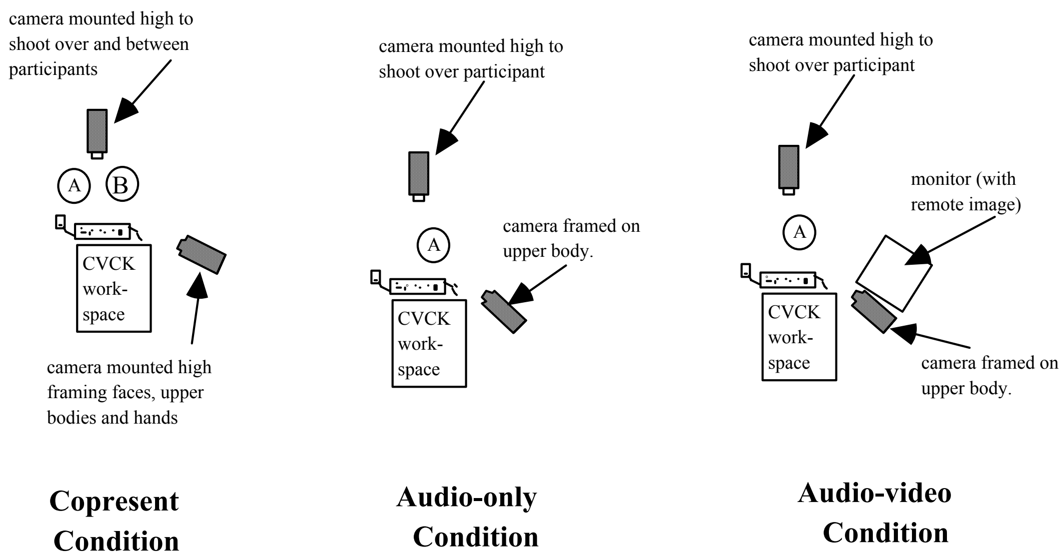


Figure 3.3: Arrangement of cameras and monitors in each of the three environments.

The placement of cameras used to capture interactions in this project is shown in Figure 3.3. Two cameras were used to record copresent interactions. One camera was positioned above and behind participants and tightly focused to record action in the electronic workspace (i.e. the CVCK). The second camera was positioned to provide an oblique frontal view that included the surface of the table, the workspace monitor, the mice, and the upper bodies of both participants. A third camera was required to capture interaction in the two distributed conditions since participants were in separate rooms. Specifically, the actions of each participant were recorded using a camera mounted to the side of the workspace, framing an image similar to the one recorded for copresent interactions. As in the copresent condition, a camera was mounted above and behind one of the participants and framed to record activity in the shared workspace.

Several important features of this arrangement of cameras and monitors with respect to interactions in the audio-video condition should be emphasized. First, the use of a second monitor to display the remote image of the other participant makes the direction of the participant's eye gaze more readily apparent to the analyst. Direction of eye gaze has been shown (Goodwin, 1981; Heath, 1986) to be an important resource for reconstructing the significance of action. In many existing systems (Dykstra-Erickson, Rudman et al., 1995; Ishii, 1990; Mantei, Baecker et al., 1991; Root, 1988) , the image of the remote participant is embedded in the main workspace screen, making it difficult to distinguish between gaze at the workspace and gaze at a partner's remote image, since no angular displacement of the head is required to redirect gaze from one to the other. The ability to discern direction of gaze is vital to participants as well (Heath, 1986; Short, Williams, & Christie, 1976) . Participants should be able to tell when their partners are looking at their remote monitors (i.e. looking "back at" their partner) versus when they are looking at the workspace. This sense of "implied eye gaze" is supported in the arrangement described above, by using a separate monitor to display the remote image, and mounting the camera between the workspace and the monitor. In this way, a remote observer sees a partner turn "towards" him or her as the partner turns to gaze at the remote image.

The use of multiple images to capture the interaction raises the challenging technical problem of merging all of these images onto a single videotape. Because the videotape will be played and replayed continuously during subsequent analysis, a video record distributed across several videotapes would be extremely impractical. This problem was solved by using video processors (see Appendix B for details) to combine all of the images onto a single videotape. For copresent interactions, the image of the workspace was inset into the image of interacting participants, placed in the lower right corner in such a way as to obscure nothing important. For the two distributed conditions, the workspace image was inset into the top left corner of one participant's image, while the other participant's image was inset into the lower left corner. Again, both insets were positioned to avoid obscuring action in the main image.

Two videotape recorders were used to simultaneously make two identical recordings of the interaction to automatically provide a set of backup videotapes. The audio from each participant's lapel microphone was recorded to a separate audio channel on the videotapes. This feature proved to be of great value during analysis, by clearly revealing which participant had spoken or even clicked the mouse based on which channel the audio was coming from.

In addition to the videotape record of the interaction, all sessions were monitored from a nearby room and extensive field notes taken.

3.3.7 Procedure

Participants were recruited in pairs as described earlier and randomly assigned to one of the three communication environments being compared. After preliminary paperwork (e.g. signing consent forms), participants were seated in front of their screen(s) showing the running CVCK simulator, and fitted with lapel microphones. In the two distributed conditions, participants were also fitted with lightweight headphones. The instructions were then reviewed by the experimenter by way of introducing the experiment:

1. The task is to follow the instructions in the laboratory manual and mark answers to questions posed therein.

2. The focus of the analysis is on collaborative interaction, not the answers to the questions; answers will not be graded for correctness.

3. The instructions in the laboratory manual are purposefully vague; participants should do the best they can. There is no time pressure.

4. The workspace cursor is controlled simultaneously by both mice.

The videotape recorders were then started and the experimenter left the room as the participants began working on the task. The experimenter electronically monitored the interaction from an adjacent room, but in no way interfered until participants indicated that they had finished.

3.4 Analysis

The videotape data for the experiment was collected over a period of about two weeks. To avoid unnecessarily perturbing the equipment, all four copresent interactions were recorded, followed by the four audio-only interactions and, finally, the four audio-video interactions. This yielded a total of 12 interactions ranging between eight and 30 minutes in length, depending on how much trouble participants experienced in interpreting and implementing the minimalist instructions given in the laboratory manual.

The following sections provide a brief overview of the analytic procedures applied to perform each of the three phases of the Breakdown Analysis undertaken in this dissertation; a more detailed discussion of these procedures is provided in subsequent chapters, which present the results of the analysis.

3.4.1 Phase 1: Identifying Consistent Patterns of Breakdown

The initial qualitative phase of Breakdown Analysis closely followed the analytic methods of Interaction Analysis to develop a preliminary understanding of the communicative breakdowns that occurred in the videotaped interactions. Videotapes were intensively reviewed to create a content log of each interaction, detailing the nature and location of episodes of communicative trouble experienced by participants; particular attention was paid to the verbal and non-verbal features of interactions in which the observed breakdowns were manifested. Based on this analysis, a notational schema was developed for textually representing the interactions, drawing on notational conventions developed in existing efforts to transcribe verbal (Atkinson & Heritage, 1984; Sacks, Schegloff et al., 1974) and non-verbal (Goodwin, 1984; Heath, 1984; Heath, 1986; Schegloff, 1984; Suchman, 1987) aspects of naturally-occurring interactions. The main features of the notational schema that resulted can be summarized as follows:

<u>1. Verbal:</u> The notations used to transcribe participants' utterances closely follow those originally developed by Gail Jefferson and used extensively by conversation analysts (Atkinson & Heritage, 1984) . Several minor modifications are introduced to take advantage of the advanced typesetting capabilities of a modern word processor.

<u>2. Non-Verbal.</u> The approach used to denote non-verbal behaviors is somewhat unusual, relying on a *landmark* model to show the relationship between verbal and non-verbal action. In essence, transcribed utterances are annotated with superscript markers that refer to descriptions of co-occurring non-verbal activities. Notation of the speaker's eye gaze is accorded special treatment by encoding it in the typeface used to transcribe a speaker's utterances.

A detailed description of the notational schema is presented in Chapter IV.

Using the notational schema developed, all of the videotaped interactions were transcribed in their entirety over a period of approximately two months, yielding a total of 346 pages of transcript for all 12 interactions.

A detailed transcript-based analysis was then undertaken to refine the categories of breakdown identified during the earlier analysis of the videotaped interaction. Strong evidentiary criteria for identifying each category of breakdown were developed and refined by iteratively evaluating them with respect to individual episodes of breakdown documented earlier.

It is important to point out a fundamental constraint on any effort to define operational criteria for identifying breakdown. The determination of whether a particular exchange does, in fact, constitute a breakdown in communication is an inherently subjective assessment based on the analyst's effort to retrospectively reconstruct the communicative significance of the behaviors documented in the transcript. Since the significance of action is intimately dependent on the unique contingencies of the local context in which it is embedded, it is impossible to define deterministic, context-independent heuristics for recognizing breakdown based either on abstract features of the interaction (e.g. timing of utterances) or on the specific content of interaction (i.e. specific phrases or words). Consequently, the criteria developed for operationalizing the categories of breakdown identified during this phase of the analysis were inevitably based on subjective heuristics for contextually evaluating the significance of action to decide whether a breakdown has occurred or not.

3.4.2 Phase 2: Statistical Assessment of Relative Communicative Efficacy

Using the operational criteria developed in the preceding analysis, the transcripts were analyzed once again to expose all episodes of breakdowns in each category that occurred over the course of each interaction, yielding a quantitative measure of communicative efficacy. Standard nonparametric statistical techniques were then applied to test for significant differences in the number of the breakdowns that occurred in interactions taking place in different environments. More formally, the Mann-Whitney U test was used to perform an analysis of variance on the number of breakdowns that occurred in various interactions. Specifically, the independent (between subjects) variables were the three communication environments, the dependent (within subjects) variables were the number of breakdowns that occurred in each category. For each statistical comparison performed, the null hypothesis was that there were no significant differences in the number of breakdowns that occurred in the environments being compared. The significant differences in amount of breakdown revealed by this analysis were used to draw conclusions regarding the overall differences in the communicative efficacy supported by each of the environments.

3.4.3 Phase 3: Rationalizing Differences

Finally, the differences in communicative efficacy exposed by the quantitative analysis were used to focus further qualitative analysis aimed at explaining those differences. For each category in which there were significant differences in frequency of breakdown between environments, breakdowns that occurred in environments that showed significantly more breakdowns (i.e. the “worse” environments) were subjected to a detailed qualitative analysis. As mentioned earlier, the aim of this analysis was to rationalize these breakdowns by highlighting constraints on certain communicative resources imposed by the environment. For example, some of the communicative resources that might be compromised by a given technologically-mediated environment include[9] eye gaze, deictic gesture, prosodic effects, body position, head movements and so on. If such resources are inaccessible to participants as they work to collaboratively establish and maintain shared interpretations of each other’s communicative displays in the given environment, their interactions will be more prone to breakdown.

To expose the causal relationships between constrained access to communicative resources and the breakdowns observed in environments with (relatively) lower communicative efficacy, all of the interactions were first surveyed to characterize the types of communicative resources that participants relied on to organize those aspects of their interaction related to each category of breakdown. That is, what communicative displays did participants produce to make available their interpretation of ongoing action to their partners and thereby *avoid* breakdown? Episodes of breakdown documented in environments with (relatively) lower communicative efficacy were then collected. For each category of breakdown in which there were significant differences between environments, the breakdowns were qualitatively examined to expose consistent patterns in the communicative resources that participants were relying on when breakdowns occurred. By showing that breakdowns that occurred in a given environment were consistently associated with the availability of certain communicative resources, the analysis strongly implies that access to these resources was somehow compromised by the environment. To further strengthen the analysis, interactions that occurred in environments with relatively higher communicative efficacy were examined to demonstrate that no similar patterns of breakdown occurred in these interactions, implying that no resource constraints existed in those environments. In the final step of the analysis, the resource constraints exposed by the analysis are related to the physical characteristics of the environment from which they arise. By establishing a causal relationship between the physical design of the environment and the communicative efficacy of interactions in that environment, the results of this analysis provide a strong basis for understanding the functional utility of the technologies and techniques used to implement current technologically-mediated environments and, ultimately, inform the design and development of future systems.

3.5 Summary: Method of Comparative Analysis

The goal of this chapter has been to introduce the method applied to yield a comparative analysis of the communicative efficacy of copresent and technologically-mediated interaction. Breakdown Analysis was introduced as a powerful analytic tool for comparing the communicative efficacy of interaction in different communication environments. Specifically, Breakdown Analysis was characterized as a four-step analysis consisting of a data collection effort, followed by three interleaved qualitative and quantitative studies. These steps are summarized as follows:

1. Data Collection: Videotape pairs of participants performing a given collaborative activity in each communication environment.

2. Qualitative Study #1: Identify and progressively refine consistent categories of breakdown through a process based on intensive review of the videotapes and ending with detailed analysis of transcribed interactions. Chapter IV presents the results of this study, discussing the patterns of communicative breakdown identified and providing extensive examples to illustrate the criteria that were developed to operationalize these categories.

3. Quantitative Study: Expose differences in communicative efficacy that exist between environments by statistically comparing the number of breakdowns documented in interactions that took place in different environments. The results of this study are presented in Chapter V, revealing significant differences in the communicative efficacy of copresent and technologically-mediated interactions.

4. Qualitative Study #2: Use differences in efficacy to focus further qualitative analysis aimed at explaining those differences in terms of resource constraints imposed by the communicative environment. Chapter VI presents the results of this qualitative study, characterizing the communicative resources used by participants to inform various aspects of their interaction, and revealing the ways in which certain resources were inaccessible in technologically-mediated interactions, rationalizing the higher incidence of communicative breakdown observed in the two distributed conditions.

Having established a firm methodological foundation, the comparative analysis of copresent and technologically-mediated environments was formally introduced by describing the method used to generate the data for the analysis. The experiment can be summarized as follows:

1. Environments: Communicative efficacy of interaction under three conditions was comparatively evaluated: copresent, audio-only, and audio-video.

2. Task: Participants used a simple cardiovascular simulator to construct and analyze the hydraulic behaviors of several cardiovascular constructs.

3. Participants: Participants had no previous exposure to technologically-mediated interaction (aside from telephonic) and had never used the simulator.

4. Experiment: The interactions of four pairs of participants were videotaped in each of the three environments, yielding a total of 12 interactions to serve as the basis for the analysis.

The following chapters present the results of applying Breakdown Analysis to comparatively evaluate the communicative efficacy of interactions in the three environments examined in this dissertation.

3.6 Notes

---

[1] In principle, the more interactions recorded for each environment the better, since conclusions will eventually be drawn from the statistical comparison of interactions in those environments. In practice, the number of interactions analyzed is constrained by pragmatic realities like the amount of time and effort that can be invested.
[2] It has been pointed out (Mackay, 1989) that ESDA is closely related in philosophy to Exploratory Data Analysis (Tukey, 1977) , though the two differ greatly methodologically.
[3] A survey of software-based shared workspaces (e.g. Timbuktu ™, In Person ™, various custom software) showed that all such applications were plagued by substantial delays, lost data, and "jerky" performance when challenged by the heavily graphical nature of the cardiovascular simulation used by participants in this experiment.
[4] In practice, this modification was relatively easy to accomplish since video images of each participant were already being recorded as part of the record of interaction (i.e. as data for the upcoming interaction analysis) in the Audio-only condition; these video images were merely redistributed to monitors in front of each participant to establish a video connection.
[5] Even in cases where both participants were apparently attempting to move the cursor to the same location, the slight path differences and compounded vector resulted in erratic or inaccurate pointing.
[6] The CVCK system was originally developed during a project funded by the Federal Institute of Post-Secondary Education (FIPSE) aimed at exploring the potential of simulation-based electronic learning environments as support for college level biology curricula. The CVCK system was recently published on CD-ROM as part of a collection of science-related learning environments by the BIOQUEST project at the University of Maryland.
[7] Though the CVCK simulator is currently used in some of these classes as part of the curriculum, none of the participants had been exposed to it at the time they participated in this experiment.
[8] This strategy is partially motivated by similar projects (Fox, 1993) in which participants were motivated by a personal stake in the interactions recorded.
[9] It is important to emphasize that the list of potential communicative resources is essentially infinite. Under the epistemology of Situated Action, *any* feature of a particular context of action can become relevant to the intepretation of the significance of that action.

CHAPTER IV

PATTERNS OF BREAKDOWN

In Chapter III, Breakdown Analysis was introduced as a technique for exposing differences in the communicative efficacy of different communication environments by comparing the amount of communicative breakdown that occurs in participant's interactions. The goal of this chapter is to present the results of the first, qualitative phase of the analysis which was aimed at identifying patterns of communicative breakdown and establishing strong operational criteria for recognizing breakdowns in each of these categories. These categories establish a basis for comparison in the quantitative analysis presented in Chapter V.

Before introducing the categories of breakdown identified in this analysis, it is important to emphasize two points made in earlier chapters which fundamentally shape the goals and expected outcomes of this phase of the Breakdown Analysis:

1. Goals of the analysis. Though the analytic procedures used in this analysis are derived directly from Interaction Analysis, the goals of Breakdown Analysis are quite different. Where Interaction Analysis is aimed at documenting the communicative regularities that participants orient to as they work to construct shared interpretations of action, Breakdown Analysis seeks to expose differences in the amount of breakdown experienced in various interactions. Accordingly, the techniques of Interaction Analysis were applied to expose patterns of breakdown that occurred, but all further analytic effort was then focused[1] on developing strong operational criteria for recognizing instances of breakdown in each category, rather than on dissecting individual breakdowns as a way of demonstrating the existence of some conversational regularity.

2. Emphasis on Interpretive Heuristics. The epistemological foundation of Situated Action fundamentally constrains any endeavor to develop objective criteria for recognizing episodes of breakdown. Because communicative breakdown is an essentially semantic phenomenon associated with participants' contextual construction of the significance of each other's communicative displays, criteria for recognizing episodes of breakdown must take the form of interpretive heuristics which are contextually applied to determine whether communicative breakdown has occurred in a particular situation. Thus, recognizing episodes of communicative breakdown necessarily requires a certain amount of inference by the analyst, introducing a subjective component into the analysis. In light of this constraint, the goals of this analysis were not to develop operational criteria that completely eliminate subjective assessment, but to minimize the scope of any inference by firmly grounding all criteria in empirical evidence.

In sum, the goals of this initial qualitative phase of Breakdown Analysis were to establish the comparative framework of the analysis by exposing consistent patterns of breakdowns, and articulating strong operational criteria for recognizing episodes of breakdown for each category.

The following section establishes the overall framework for recognizing breakdown used in this analysis and briefly introduces the four categories of breakdown that were identified. Section 4.2 describes how the videotaped interactions were transcribed and presents a detailed discussion of the notional conventions used. Subsequent sections then discuss each of the four categories of breakdown in detail, and describe the operational criteria developed to recognize breakdowns in each category.

4.1 A Framework for Recognizing Communicative Breakdown

In Chapter II, the notion of "communicative interaction" was characterized as a collaborative process in which each participant makes available evidence of his or her interpretation of the evolving interaction through verbal and nonverbal communicative displays like utterances, prosodic effects, direction of gaze, gestures and so on, while simultaneously monitoring the reciprocal displays of conversational partners. To organize this interpretive process, participants rely on a set of socially-defined conversational regularities which embody common patterns of action and interpretation.

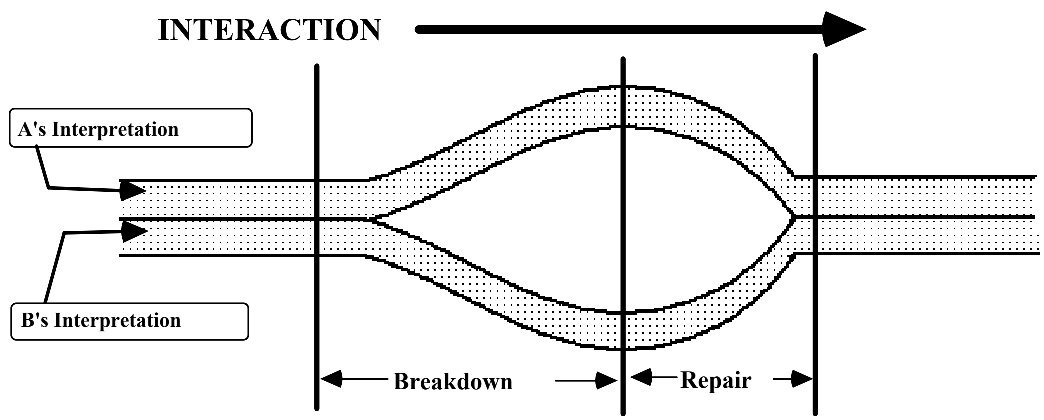


Figure 4.1: Graphical representation of communicative breakdown and repair.

As indicated in Figure 4.1, communicative breakdown is characterized as the failure of the evidentiary processes through which participants maintain intersubjectivity, resulting in a divergence of their interpretations of ongoing action and, ultimately, in communicative confusion. In most cases, this divergence is eventually detected by participants and repaired either tacitly, with the breakdown itself serving as a resource for resynchronizing participants' interpretations of action, or through a verbal repair sequence in which the confusion is explicitly recognized and collaboratively resolved.

While this abstract characterization provides a good conceptual foundation for understanding the notion of communicative breakdown, it provides no practical basis for actually recognizing episodes of breakdown in naturally occurring interactions. In particular, it fails to articulate specifically what aspects of interaction participants must organize in order for a communicative exchange to be intelligible. That is, what exactly are the organizational processes that are subject to breakdown? In short, if we expect to identify communicative breakdowns that occurred in the videotaped interaction, we need a more specific idea of what we are looking for.

A basic analytic framework[2] for the analysis was established by recognizing three fundamental organizational issues that must be continuously addressed by participants in any interaction in order for the interaction to be coherent:

1. Turntaking: Whose turn is it to contribute to the interaction? In the process of communicating, participants rely on a variety of communicative resources like shared talk, gesture, body position, and manipulation of the physical world to make available their moment-by-moment interpretations of the evolving interaction and, in this way, maintain shared interpretations of action. Because access to some of these resources is mutually exclusive, allowing only one participant or another to utilize a resource at any given moment to produce a contribution to the interaction, participants must regulate access to such mutually exclusive resources by developing some sort of turntaking systematics for passing control over the resource back and forth between them. The most obvious example of a mutually exclusive resource is the verbal channel — participants must regulate access to the verbal floor in order to avoid overlapping talk. Other mutually exclusive resources may exist in certain conversational contexts as well. This is particularly true of task-oriented interactions in which participants are manipulating the real world in some way. For example, if two architects are collaboratively sketching the plans for a house on a chalkboard, then their piece of chalk represents a mutually exclusive resource; the architects must somehow pass control over this resource back and

forth between them. In the task-oriented interactions examined in this study, the shared cursor essentially played a similar role as the chalk, allowing participants to manipulate the electronic problem representation (i.e. the CVCK simulator). Despite the fact that each participant was given a mouse, the shared cursor was very much a mutually exclusive resource since simultaneous efforts to control the cursor inevitably resulted in erratic behavior of the shared cursor. Consequently, participants had to take turns at controlling the shared cursor.

2. Topic: What are we currently talking about? The notion of topic constitutes a fundamental organizational mechanism for conversation, establishing a basis for defining the notion of "progress" in conversation. Participants refer to the topical framework of the conversation to make apparent the relationship of their current utterances and actions to previous discussion, thereby establishing the logical coherence of the conversation as a whole. For example, as the two architects mentioned earlier discuss the plans for a house, their topic of discussion may shift from discussion of the houseplan as a whole through an entire hierarchy of subtopics like where to place the bedrooms, how many sinks to use in each bathroom, and so on. Because the significance of each subtopic arises only in relation to the overall topical structure within which it is embedded, it is vital for participants to maintain synchronous conceptions of the current topic — what it is that their talk is currently about — as the interaction evolves.

3. Reference: What are you referring to? In order for conversation to be intelligible, participants must continually establish shared reference to the objects and entities that exist in the current context. For example, in order for our architects to construct shared interpretations of the utterance "Let's move the carport over next to the guest bedroom," they must established a shared understanding of what graphical components of the houseplan they are working on are being referred to as the "carport"

and the “bedroom.” In negotiating shared reference, the listener examines the current context for an appropriate referent, while speaker monitors the listener’s verbal and nonverbal displays for evidence that the referent has been unambiguously located.

The conversational resources and techniques used by participants to manage each of these three organizational activities has been extensively examined in existing research.

Beginning with Sacks’ (1992a; 1992b; 1992c; 1974) seminal work on verbal turntaking behavior, a number of conversation analytic studies[3] (Pomerantz, 1978; Schegloff, 1987; Schegloff, Jefferson et al., 1977) have examined the verbal mechanisms that participants rely on to organize their turns at talk. More recently, several interaction analytic studies (Fox, 1993; Heath, 1984; Heath, 1986; Schegloff, 1984) have expanded this investigation to consider the role of nonverbal resources like eye gaze and body position in regulating access to the verbal channel. Social psychologists have studied verbal turntaking as well, albeit from a more mechanistic perspective. For instance, several studies have worked to characterize the role of eye gaze (Champness, 1970; Kendon,

1967) and various verbal "signals" (Argyle, 1969; Cook & Lalljee, 1972) in negotiating turns at talk.

The issue of how conversational participants maintain shared topical orientations while progressing from one topic to the next over the course of interaction has received extensive attention in the literature as well, both with respect to verbal mechanisms for negotiating topic openings and closings (Beach, 1990; Beach, 1993; Button & Casey, 1984; Covelli & Murray, 1980; Erickson, 1982; Jefferson, 1993; West & Garcia, 1988) and nonverbal displays used by participants to display their current topical orientation (Fox, 1993; Heath, 1986).

Finally, a number of studies have worked to articulate the verbal (Anderson, Bader, Bard, Boyle, Doherty, Garrod et al., 1991; Clark & Brennan, 1991; Clark & Schaefer, 1989; Clark & Wilkes-Gibbs, 1986; Isaacs & Clark, 1987; Wilkes-Gibbs & Clark, 1992) mechanisms that participants use to establish and maintain shared reference. Several interaction analytic studies (Fox, 1993; Heath, 1984; Heath, 1986) have focused on the nonverbal displays used by participants to negotiate reference as well.

Importantly, these three fundamental organizational activities — organizing turntaking, organizing topic, and organizing reference — provide a concrete framework for understanding what it means for "breakdown" to occur in interaction — breakdowns occur when participants' organizational efforts fail, resulting in divergent conceptions of either what the current topic is, what object or entity is the referent of an immediately preceding utterance, or whose turn it is to contribute to the conversation. The resulting framework for recognizing breakdowns in the interaction is depicted in Figure 4.2:

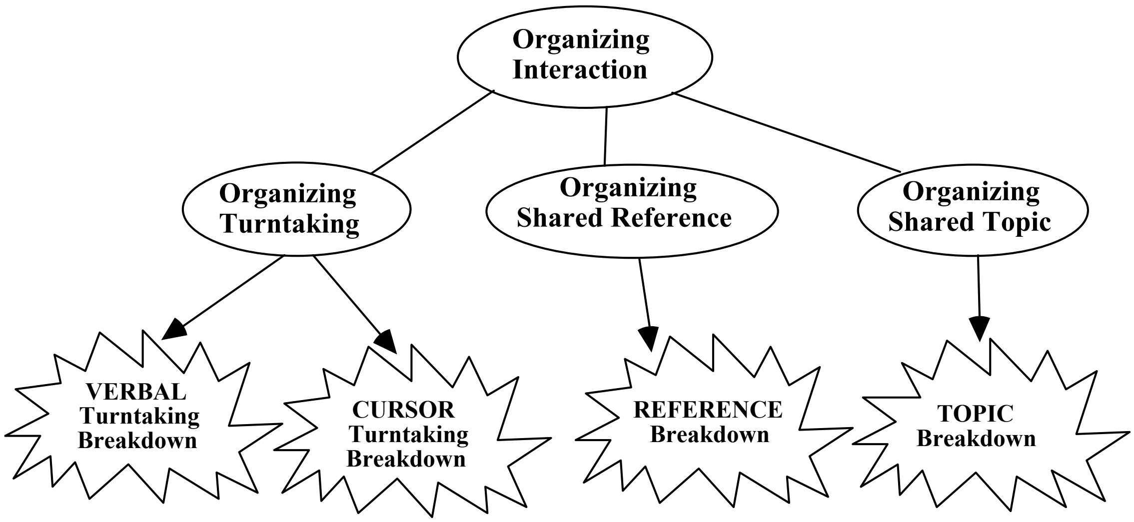


Figure 4.2: A framework for recognizing breakdown in interaction.

As indicated in Figure 4.2, four distinct categories of breakdown arise from failures to successfully manage the three organizational activities described earlier; the following sections present an overview of each category.

4.1.1 Verbal and Cursor Turntaking Breakdown

Both Verbal and Cursor turntaking breakdowns are defined by failures to organize turntaking during the interaction, resulting in confusion over whose turn it is to utilize a certain resource to contribute to the interaction. In the case of Verbal turntaking breakdown, the resource in question is the verbal channel; Cursor turntaking breakdowns are defined by failures to regulate access to the shared cursor.

For both Verbal and Cursor turntaking breakdowns, recognizing that a breakdown has occurred is centered around finding evidence that participants' conceptions of who currently controls the communicative resource in question (i.e. either the verbal floor or the shared cursor) have diverged. In general, there are two possible ways in which such divergent interpretation of who controls a resource can occur, each of which is evidenced differently in the record of interaction:

i. Each participants simultaneously believes that it is his or her turn to control the resource in question. This condition is evidenced by an attempt by both participants to simultaneously utilize the resource.

ii. Participants both believe that it is a partner's turn to control the resource in question. This condition is evidenced by a period of inactivity during which each participant waits for his or her partner to act.

Clearly, the second of these two conditions is much more difficult to unambiguously recognize than the first, since it requires somehow distinguishing periods of inactivity that arise from turntaking confusion from mundane pauses in interaction. Even clear evidence of a simultaneous attempt to utilize a resource does not necessarily indicate confusion (i.e. breakdown) over whose turn it is to control the resource. In particular, it may be the case that a participant is *willfully* interrupting a partner's turn at controlling the resource in question. The operational criteria developed to recognize both Verbal and Cursor turntaking breakdowns in sections 4.3 and 4.4 explicitly address these issues.

4.1.2 Topic Breakdown

Topic breakdowns are defined by a failure of the conversational mechanisms participants rely on to maintain shared topical orientations, resulting in a situation in which participants have different conceptions of the current topic of conversation. For example, if two participants are working on collaboratively answering a series of questions, a Topic breakdown has occurred if one participant believes that a question has been answered and that discussion has moved to the next question, while the other participant still believes the discussion to be focused on the previous question.

Because participants' beliefs regarding the current topic of discussion are not directly related to unambiguous physical behaviors, the only way to reliably detect Topic breakdowns is through explicit evidence in the verbal channel that participants' topical orientation has diverged.

4.1.3 Reference Breakdown

Reference breakdowns are defined by failures of participants to establish shared reference to the objects and entities in the referential context. Specifically, a Reference breakdown has occurred when either the speaker or the listener becomes uncertain that a linguistic reference produced by the speaker in an immediately preceding utterance has been understood by the listener.

Like confusions about current topic, Reference breakdowns are not necessarily manifested in the observable physical behaviors of participants. Consequently, the operational criteria developed for recognizing Reference breakdowns are centered around explicit verbal repair sequences initiated by either participant, in which the referential confusion is made apparent to both participants and collaboratively resolved.

4.1.4 Summary

In sum, a set of *analytic foci* for guiding the analysis of the videotaped interactions can be established by recognizing that participants in any conversation must continuously attend to three organizational tasks in order to maintain mutual intelligibility. First, participants must somehow organize access to certain mutually exclusive communicative resources, establishing turntaking mechanisms to pass control over those resources back and forth between them. Though the verbal channel represents the most obvious and ubiquitous example of a mutually exclusive resource, manipulative tools like the shared cursor used by participants in the CVCK task also constitute mutually exclusive communicative resources. Second, participants must maintain shared conceptions of current topic, making available their moment-by-moment topical orientations while monitoring the communicative displays of partners for evidence of their conceptions of current topic. Finally, participants must maintain shared reference to the objects and entities in their conversational context, monitoring their mutual interpretations of the linguistic references made by a speaker.

Distinct categories of communicative breakdown can be defined in terms of the failure of each of these organizational processes, resulting in divergent interpretations of current topic, direct references, or control over either the verbal floor or the shared cursor.

The following section describes how the analytic framework established in this section was used to guide the development of a notational schema for transcribing the videotaped interactions. This sets the stage for a more in-depth discussion of each of the four categories of breakdown and the description of the evidentiary criteria developed to operationalize each category presented in subsequent sections.

4.2 Creating Transcripts of Interaction

A detailed textual transcript of interaction is a vital resource for any analytic technique based on the in-depth examination of communicative interaction. In the context of Breakdown Analysis, the textual transcript provides the foundation for all later stages of the analysis, from the operationalization of the patterns of breakdown identified during preliminary analysis of videotapes, to the final stages of the analysis in which individual episodes of breakdown are examined to expose constraints on communicative resources imposed by technologically-mediated environments.

Clearly, it is vital for any transcript of interaction to textually represent the verbal and non-verbal behaviors of participants at a level of detail that allows this textual record to serve as the basis for further analysis of the patterns of behavior that the analyst is interested in. Unfortunately, what features of interaction are "relevant" to the upcoming analysis can not be entirely known in advance. This exposes a subtle dilemma of transcription: how can one decide what features of interaction should be captured in a transcript if it is the upcoming (transcript-based) analysis that will ultimately determine which features of interaction are relevant for evidencing the patterns of behavior of interest to the analyst? This circular dependency is emphasized by Jordan and Henderson (1995) in their treatise on Interaction Analysis:

> It makes sense, then, for researchers to think very seriously about what kind of analysis they intend to do before launching into full-scale transcription, because the choice of what to transcribe determines what will be available for analysis. ... Nonetheless, it is impossible to include *all* potentially relevant aspects of an interaction, so that, in practice, the transcript emerges as an iteratively modified document that increasingly reflects the categories the analyst has found relevant to her or his analysis. (Jordan & Henderson, 1995, p. 10, original emphasis)

As indicated by Jordan and Henderson, the only solution is to intertwine transcription and analysis to some extent, allowing the evolving understanding of the interactions documented on the videotape to shape the representational priorities of the notational schema developed for the transcription.

Accordingly, transcription of the interactions analyzed in this study took place only after an intensive review of the videotape data aimed at understanding how breakdowns in the four categories described earlier were manifested in the observable behaviors of participants. That is, what verbal and nonverbal behaviors available in the videotape records of interaction are vital for recognizing episodes of Verbal turntaking, Cursor turntaking, Reference, and Topic breakdown? Based on this preliminary analysis of the videotape data, a notational schema was developed, drawing on existing notational conventions developed to denote verbal (Atkinson & Heritage, 1984; Sacks, Schegloff et al., 1974) and non-verbal (Goodwin, 1984; Heath, 1984; Heath, 1986; Schegloff, 1984; Suchman, 1987) aspects of interaction. Table 4.1 summarizes the notational biases of the schema.

Table 4.1: Summary of representational priorities of the notational schema developed for the analysis.

High fidelity	Low fidelity	Ignored
Temporal Relationships	Prosodic Effects	Auto-manipulation
Utterances	Body Posture	Fidgeting
Direction of Gaze	Facial Expression	Incidental sounds
Cursor Movement	CVCK behavior	
Hand/Mouse Movement	Peripheral sound	
Hand Gesture/Pointing		

As indicated in Table 4.1, a strong representational focus was placed on capturing the nonverbal behaviors of participants with respect to the workspace and laboratory manual, and the context of the utterances that accompanied them. For example, the temporal relationship between utterances and nonverbal events were rigorously represented, as were the movements of the shared cursor and participants' hands and direction of gaze. More subtle characteristics of the interaction like facial expressions, prosodic effects, and body posture were often represented as well, though at a generally lower level of fidelity. For example, facial expressions were denoted only when they were extremely overt or seemed particularly relevant to ongoing events; body posture was denoted only when large scale changes in body position occurred. Finally, most auto-manipulative behaviors like scratching, brushing hair from the eyes, and adjusting clothing were completely ignored, as were incidental sounds like distant slamming of doors, squeaking of chairs and so on.

4.2.1 Overview of Transcription Format

Although the medium and format of transcripts is changing rapidly (Mackay, 1989; Psathas & Anderson, 1990) , textual transcripts have traditionally been generated by a standard typewriter, limiting notational schemes to plain (i.e. single font, single font size, no emphasis) text representations. The notational system described here takes advantage of the type-setting conveniences afforded by modern electronic word processors to simplify the notational system and streamline the transcription process. The overall pragmatic constraint used to guide the development of the notational system was that the transcript must be printable on standard-sized sheets of paper using an easily readable font size.

The following segment of transcript illustrates the basic format of the notational schema used in this work:

	Verbal	Non-verbal	CVCK
•	R: *then it says to convert the gauges into graphs by doubleclicking on* <u>*each*</u> of them ((3)) R: well --- lets see (1.2) gotta [4]doubleclick on that[5] and (.6) [6]doubleclick on that --- hey!	3- R rolls cursor to V1 and pauses, M stares LB (2.9) 4- R doubleclicks V1 5- M raises to WS 6- R doubleclicks on V2	4- parameter dialog for V pops up 6- CVCK boings twice because of open dialog

Segment: AV3p24

The format essentially consists of three[4] columns that document, respectively, the events in the verbal channel, the non-verbal behaviors of participants, and the behaviors of the CVCK simulator. In addition, the small leftmost column provides a way of drawing the reader's attention to relevant utterances or actions within an exchange, by placing a small black dot (•) adjacent to them. For example, when segments of transcript are used in upcoming sections to illustrate the criteria developed for recognizing breakdowns in the four categories mentioned earlier, a black dot in the leftmost column is used to identify the point in an exchange at which a breakdown occurs.

Segment names refer to the interaction from which they are drawn. The first two characters indicate the environment in which the interaction occurred: FF = copresent (face-to-face) environment; AO = audio-only environment; AV = audio-video environment. Remaining characters denote which interaction the exchange is drawn from, and the page of the transcript on which it occurs. Thus, Segment AV3p24 appears on the 24th page of the audio-video interaction labeled AV3.

The following sections describe the various notational features evident in Segment AV3p24 in more detail.

4.2.2 Representing Temporal Aspects of Interaction

One feature of real interactions that is easily lost in any transcription effort is the temporal aspects of the interaction. Questions include "How long did it take to produce that utterance?," "What was the elapsed time between those two utterances?" and "How long did that non-verbal activity go on?" Short of somehow placing all utterances and events on some sort of continuous timeline running through the transcript, it is very difficult to accurately preserve these temporal features. As a compromise, a concerted effort was made to time all silences between utterances, including non-verbal activities that occurred during such silences. However, the temporal extent of utterances themselves was not explicitly timed and must be inferred from the textual representation of the utterance.

All timing was performed manually using a standard stopwatch; elapsed times are shown to the nearest tenth of a second.

4.2.3 Documenting Nonverbal Behaviors

The primary issue to be addressed by any notational schema that works to capture both verbal and non-verbal aspects of interaction is how to denote the temporal relationship between verbal and non-verbal events. That is, how accurately does the notation represent the relationship of non-verbal behaviors to the utterances being produced in the conversation? At the same time, the notation must conform to certain pragmatic constraints. For instance, the resulting schema must be relatively compact and can not be so dense or complicated that it becomes unreadable.

The spectrum of possibilities for denoting the relationship between verbal and nonverbal behaviors is defined at the one end by approaches (e.g. Suchman, 1987) that only loosely correlate non-verbal behaviors with co-occurring utterances and, at the other, by approaches (e.g. Heath, 1986) that painstakingly document the moment-by-moment relationship between verbal and non-verbal aspects of interaction. The approach taken in this work falls roughly in between these two extremes, adopting a *landmark* model[5] for denoting non-verbal behaviors, in which the verbal record is annotated with markers that denote the occurrence of some non-verbal event. In this way, the notation precisely preserves *when* a non-verbal event was initiated, but only loosely documents the *extent* of that event. The "landmarks" in the notation consist of numerical indices that refer to descriptions of non-verbal behavior that are separate from the verbal transcript. Two forms of "landmarks" are used to annotate the verbal transcript to reflect non-verbal events:

1. Superscript Indices. If the non-verbal event occurs while an utterance is being produced by one of the participants, a superscript marker is placed in the verbal transcript at the point at which the non-verbal behavior is initiated. For example, the transcript shown in Segment AV3p24 indicates that participant R doubleclicks her mouse with the cursor centered on component V2 just as she says "gotta doubleclick". In the rightmost column, we can see that the CVCK responds to this action by popping up a dialog box.

2. Parenthetical Indices. To document non-verbal events that occur during periods of silence, the numerical index to the appropriate descriptions of non-verbal events is enclosed in double parentheses. For example, the "((3))" in the first column of Segment AV3p24 indexes a description of non-verbal events that occurred during the silence between the two utterances shown. Note that temporal extent of silent nonverbal events is noted as well, with the total time of the event appearing in single parentheses following its textual description in the Non-verbal column of the transcript.

By adopting a landmark approach to documenting participants' non-verbal behaviors, the notational schema provides a reasonable amount of temporal detail, while avoiding excessive notational clutter that would only obscure the analysis.

4.2.3.1 Direction of Eye Gaze

Another important feature of the notational schema demonstrated in Segment AV3p24 is the special attention accorded to documenting the speaker's direction of gaze. In general, three distinct directions of gaze were recognized in the notation: gaze directed at the shared workspace, gaze directed at the laboratory manual, and gaze directed at the other participant[6]. The moment-by-moment direction of the speaker's gaze is denoted by the typeface used to transcribe his or her utterances — plain typeface indicates gaze directed towards the workspace, italics typeface indicates gaze directed at the laboratory manual, and bold typeface indicates gaze directed at the other participant. It should be emphasized that the advantages of this approach are purely practical. That is, changes in gaze by a silent participant are just as accurately denoted using the landmark approach described earlier; the direction of a silent participant's gaze at any moment in the interaction can be easily determined based on this information. However, the use of different typefaces makes the speaker's direction of gaze more readily apparent to the reader and substantially reduces the number of landmarks (i.e. clutter) inserted in the verbal transcript.

When changes in the direction of a silent participant's eye gaze occur, the are denoted in the same way as all other non-verbal events, by using the landmark approach described earlier. That is, a numerical index is inserted into the verbal transcript and the change in eye gaze is described in the Non-Verbal column of the transcript. For compactness, the three directions of eye gaze that were distinguished are abbreviated as follows: WS = workspace; LB = laboratory manual; RS = the remote participant. For example, annotation number three in Segment AV3p24 reads "R rolls cursor to V1 and pauses, M stares LB (2.9)". This annotation indicates that, during the 2.9 second silence separating the two utterances shown in the segment, participant M gazed at the laboratory manual while participant R rolled the cursor to component V1 in the workspace.

4.2.3.2 Referring to Components of the CVCK

A particularly important class of non-verbal behaviors is participants' manipulation of the CVCK simulator and the responses of the simulator to those actions. Denoting these behaviors mainly involves textually describing participants' manipulations of the iconic components that make up the simulator interface — which cardiovascular component they are dragging, what they are clicking on using the shared cursor, and so on.

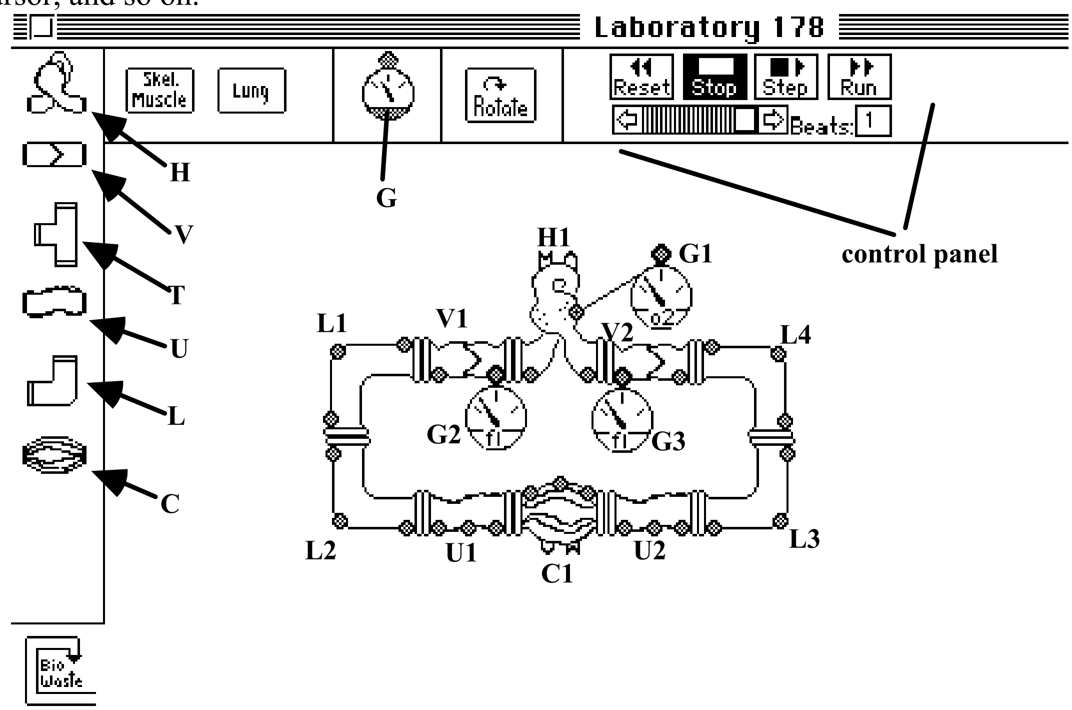


Figure 4.3: Labels used to refer to various components of the construction created by participants.

Because these interface components are, for the most part, not textually labeled in the CVCK interface, a convenient labeling system was developed to allow components to be easily identified in the transcript. Figure 4.3 shows an annotated image of the simulator workspace in which all of the components of the construction have been labeled.

Those few components that do have textual labels incorporated in their iconic representations (e.g. biowaste, run button) are referred to by those labels. All other components were referred to by the labels shown in Figure 4.3. For example, the description “R doubleclicks on V2” given in the Nonverbal column of Segment AV3p24 indicates that participants R clicked on the rightmost valved vessel in the construction.

It is important to point out that these labels refer more to positions in the construction than to components themselves. What this means is that the label used to refer to a single component in the transcript could change over the course of the construction process, as it is placed in various positions by participants. For example, as a new ventricle (H) is dragged from the palette, it is referred to as “H”. As soon as that piece is positioned and placed in the proper place in the evolving construction, however, it becomes “H1”. Another condition under which a component’s label in the transcript might change is when it is erroneously placed in the wrong position in the construction. For example, suppose[7] that participants erroneously install an unvalved vessel (U) in place of the valved vessel V1. This component would be referred to as “V1u”, indicating the that it is an unvalved vessel (U) erroneously placed as V1. When participants discover the mistake and correctly reposition the unvalved vessel in the lower half of the construction, its label would be changed accordingly (i.e. to either U1 or U2).

4.2.4 Transcribing Verbal Behavior

In this section, we turn to a detailed overview of the notational symbols that appear in the Verbal column of the transcript notation. The transcription system used for talk is adapted directly from the one developed by Gail Jefferson, which can be found in (Atkinson & Heritage, 1984) . Table 4.2 summarizes the notation conventions used in the Verbal column of the transcript.

4.2.5 Summary: Creating Transcripts of Interaction

Using the four categories of breakdown established earlier to focus attention on certain organizational behaviors, the videotape data was carefully reviewed, paying particular attention to the observable features of interaction through which episodes of breakdown are revealed to the analyst. Based on this analysis, all of the interactions were transcribed in their entirety, carefully preserving verbal and non-verbal features of the interactions relevant to evidencing the communicative difficulties identified. These transcripts then served as the basis for a more detailed analysis aimed at refining the observed patterns of breakdown and, in particular, developing strong operational criteria for recognizing breakdowns in each category.

The following sections provide a detailed discussion of the evidentiary criteria developed to recognize each of the four categories of breakdown examined in this study.

Table 4.2: Summary of notational conventions used in the Verbal column of transcripts.

Symbol	Example	Explanation
[	R: ohhh okay::: [M: I guess::::	A left bracket indicates the point at which a speaker's utterance is overlapped by a partner's talk. In keeping with the landmark approach, the extent of the overlap (i.e. when it ends) is implied by talk, rather than being explicitly denoted.
=	R: *youcn go ahead and take* that thing thats-= M: =*okay:: - can I::: --- start from:::::*	Matching equals signs, one at the end of a line and the other at the beginning of the next indicate that there was no gap between the utterances.
(value)	R: *the* middle or whatever (.9) M: uh-- *the center?*	Values in single parentheses indicate the elapsed time between utterances. Utterances not explicitly separated by values in parentheses can be assumed to have a gap of less than 0.3 seconds between them (i.e. too small to time).
---	M: *so we haveta just* (1.0) *(dontcha) try to do*[4] *this ---- figure one::?* (.4)	Pauses within an utterance of less than half a second are often represented as a series of dashes, with each dash representing approximately one tenth of a second.
:::	M: so::: - dyou wanna do the next? R: ye::a::h	Colons indicate a lengthening of the immediately preceding sound. The number of colons reflects the amount of prolongation.
?	M: uuum --- can I do that? R: sure- dyou kno+ - ohhh theres a rotate (.5) up there	Question marks are used to indicate a rising intonation. They are not used as punctuation.
!	R: oop! ---- I dropped it- uhhhuhuhu (come on) M: aaooh	Exclamation marks denote an excited or animated tone of voice.
.	M: oops R: uhhhheyeah I dont know what happened.	A period indicates a full stop with falling intonation.

Table 4.2 (continued)

Symbol	Example	Explanation
,	M: ahh good ([4].9) *ahhhh, next one::?* R: [5]mhmmm	Commas are indicate a boundary between intonations. For instance, in the example given, the comma marks the boundary between the steady intonation of the first half of the utterance, and the rising intonation at the end.
ALL CAPS	R: ohh I see::- ther::e- look [3]at that thing= [M: (u uh) R: =oh GOO::D	Capital letters indicate that the capitalized sounds are much louder than surrounding talk.
smaller font size	M: lung::: -------- skeletal muscle [[R: uhhu-h-h huhi doennoe	A smaller font size is used to indicate that the utterance is much quieter than surrounding talk. The smaller the font size, the closer the utterance is to a whisper.
{text}	(.6) R: {cough} [M: on::::, bee now?	Curly braces are used to indicate sound effects, throat clearing, coughing, or other sound effects that are not easily described phonetically.
(text)	M: *so we haveta just* (1.0) *(dontcha) try to do*[4] *this ---- figure one::?* (.4)	Text in parenthesis indicate transcriber's uncertainty as to the utterance. The text represents the best guess or, when utterances are totally incomprehensible, a phonetic reproduction of audible sounds.
+	(.6) M: unkay::: on the[3] c+ --- un::: the macintosh, right? R: ri::ght	Plus signs at the end of a word fragment indicate that the word was abruptly aborted.
((text))	R: uh maybe I should reset it, huh? (.4) M: mmmm ((nods))	Text enclosed in double parentheses contain transcribers annotations to the transcription. They are often used for trivial non-verbal behaviors that don't warrant explicit description using the superscript notation described earlier.

Table 4.2 (continued)

Symbol	Example	Explanation
((integer))	R: so should we run it again? ((3)) M: mhmm ((4))	Integers in double parentheses refer to descriptions of the non-verbal behaviors occurring in the gaps between utterances given in the Non-verbal column of the transcript format.
super^{1}script	M: *ung*^{2}*--kay ----- modify the system:: (yourrsh)* (1.9) R: *attach a pressure gauge to heart* (.6) *at* point	A superscript integer refers to non-verbal behaviors described in the Non-verbal column of the transcript notation that were initiated at that point in the utterance.
.hhh	M: .hhh yeah maybe those're:::: ----- maybe:: [R: *huhu* ------ *hu*n::kay ----*hehe*	A period followed by a series of h's indicates an audible inbreath. The number of h's indicates the length of the inhalation.
hhhh	R: *.hhhh-hhhh --- convert the gauges into graphs by doubleclicking on each of them ----- move them:: so they are aligned vertically::: one above the other::*	A series of h's NOT preceded by a period indicates an audible exhalation. The length of the exhalation is indicated by the number of h's.
hahaha huhuhu hehehe uhhhuhu	M: uuuuh no:::-- uhhhuhe ------ .hhhh [R: no:: -hehehehehe (.8)	Laughter is transcribed phonetically, characterized by an alternation of h's and the appropriate vowel to describe the sound. Note that a series of h's may appear embedded in laughter, indicating laughter combined with exhalation.
h-h-h-h	(.7) R: so do you have any answers?-h-h-h-huhu-h= M: =uuuuuh	Colorless, breathy laughter that contains no vowel sounds is denoted by h's separated by dashes.
plain typeface	M: yeah::: and also the-- pre+ -- pressure graph? [R: oooh -- maybe by	Plain typeface indicates that the speaker is gazing at the workspace at the time the transcribed sounds were produced.

Table 4.2 (continued)

Symbol	Example	Explanation
bold typeface	R: this:: thing right here? ((3)) M: **I cant see where you're pointing:: you retard**	Bold typeface indicates that the speaker is gazing directly at the other participant while speaking. In the audio-video condition, gaze is directed at the remote monitor.
italics typeface	R: *I dunnoe --- I dunnoe where the heart is- uhhh-h-ha*n I dunno+ ---- I mean we dont have numbers, right?	Italics typeface indicates the speaker is gazing at the laboratory manual at the time the transcribed sounds were produced.

4.3 Operationalizing Verbal Turntaking Breakdown

Verbal productions are clearly the primary resource that participants rely on to construct shared interpretations of action; anyone who has ever tried to conduct a conversation without the benefit of spoken language can attest to the fact that achieving shared understanding under such conditions is extremely difficult. The verbal channel is not a mutually exclusive communicative resource in the strictest sense, since it is possible for several participants to speak simultaneously without rendering their utterances entirely unintelligible. However, there is extensive evidence (Sacks, 1992b; Sacks, Schegloff et al., 1974) that conversational participants work to organize mutually exclusive access to the verbal channel, relying on highly-refined conversational mechanisms for tacitly passing control of the verbal floor back and forth over the course of interaction. Verbal turntaking breakdowns are defined by the failure of these conversational mechanisms, resulting in confusion over whose turn it is to speak. The two possible manifestations of such turntaking confusion are graphically illustrated in Figure 4.4.

As indicated in Figure 4.4, participants initially have synchronous interpretations of whose turn it is to speak at each moment during the interaction, as control of the verbal floor is passed back and forth between them.

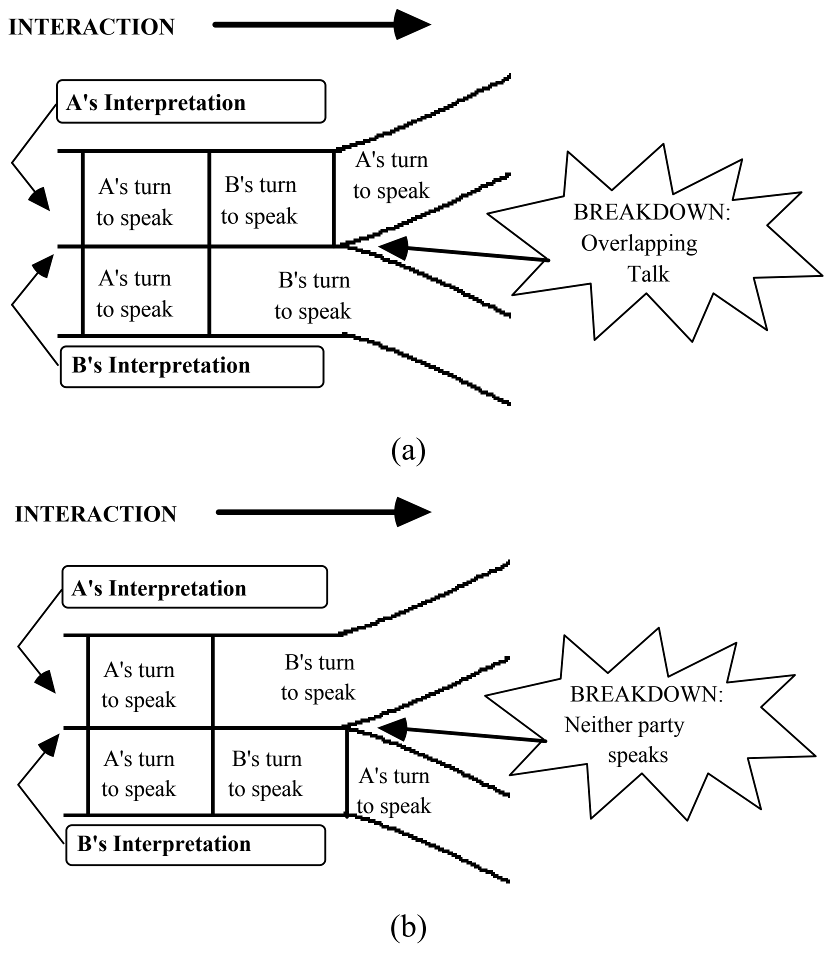


Figure 4.4: Verbal turntaking breakdown characterized as divergent conceptions of whose turn it is to speak. Either (a) both participants come to believe it is their turn to speak or (b) they each believe it is the other's turn to speak.

One way in which this turntaking process can break down (Figure 4.4b) is when a speaker ends a turn at talk and displays evidence that he or she believes that control of the verbal floor has passed to a conversational partner, while his or her partner fails to recognize the turn ending and displays evidence that he or she continues to believe the speaker to retain control of the conversational floor. Specifically, this confusion is evidenced by an extended period of silence as each participant waits for the other to speak. Unfortunately, it is impossible to unambiguously recognize this condition based on the evidence available in the record of interaction, since such confusion is almost never explicitly repaired by participants. Rather, repair is tacitly accomplished when, after some period of time, one participant simply continues the conversation by producing the next utterance. This makes it impossible to distinguish silences due to confusion over whose turn it is to speak from ordinary pauses in the conversation as participants manipulate the simulator or otherwise attend to the task they are engaged in. Accordingly, no effort was invested in recognizing this variation of Verbal turntaking breakdown.

Another manifestation of Verbal turntaking breakdown is (Figure 4.4a) when a listening participant displays evidence that he or she has come to believe that the speaker has ended an immediately preceding turn at talk and has passed control of the verbal floor to the listener when, in fact, the speaker's subsequent actions imply the he or she still believes to be in control of the verbal floor. This confusion results in an attempt by both participants to simultaneously utilize the verbal channel, evidenced by overlapping talk in the record of interaction. This characterization of Verbal turntaking breakdown as the divergent interpretation of a turn transition opportunity resulting in a condition in which participants both simultaneously believe it is their turn to speak is embodied in the following evidentiary criterion:

> <u>Criterion:</u> Verbal turntaking breakdown is evidenced by overlapping utterances immediately following a turn transition opportunity.

While this criterion focuses analytic attention tightly on instances of overlapping talk that appear in the transcript, there are two issues that must be clarified in order for this criterion to become an operational heuristic for recognizing Verbal turntaking breakdowns:

1. What sorts of behaviors constitute turn transition opportunities?

2. What does it mean for overlapping talk to occur "immediately following" a turn transition opportunity? How long after a preceding turn transition opportunity can overlapping talk occur and still be considered evidence of Verbal turntaking breakdown?

These issues emphasize the point made earlier that recognizing communicative breakdown necessarily requires a contextual interpretation of the significance of action (i.e. does it evidence underlying confusion?), and can not be based solely on context-independent characteristics of an exchange (e.g. overlapping talk). The following sections describe how each of these issues was addressed in this analysis.

4.3.1 Turn Transition Opportunities

In a bygone era of radio communication, conversational participants were expected to explicitly pass control of the verbal channel back and forth by using the word "over" to mark the end of each verbal turn. Under this explicit turntaking system, there is never any ambiguity about when it is appropriate for a listener to take over as the next speaker. By contrast, the conversational regularities that participants rely on to regulate access to the verbal floor in everyday talk (Sacks, Schegloff et al., 1974) are based on the idea that turns at talk are tacitly negotiated, with listeners continuously examining a speaker's productions for evidence that he or she has finished the turn at talk, presenting a *turn transition opportunity* that allows a new speaker to take control of the verbal floor. Verbal turntaking breakdowns occur when participants somehow fail to mutually orient to these regularities, causing a listener to erroneously conclude that the a speaker's turn has ended. Three patterns of behavior were recognized as turn transition opportunities in this study, based on conversational regularities documented in existing analyses (Sacks, Schegloff et al., 1974) of verbal turntaking behavior: extended silences, question-answer sequences, and the use of turn ending markers.

4.3.1.1 Silence misinterpreted as turn ending

Perhaps the most obvious indication that a speaker has finished speaking and that a turn transition opportunity is at hand is when a speaker's utterance ends and a certain amount of silence accrues. Verbal turntaking breakdowns result when this silence is erroneously interpreted as a turn transition opportunity by the passive participant, as illustrated in the following example:

•	(1.6) M: *kay when blood flows (.5) through a va:::lve* [R: *is it* open or clos[3]ed M: is it open or closed R: uuum (1.8)	3- M snaps to WS and grabs mouse	

Segment: AV2p9

In segment AV2p9, a Verbal turntaking breakdown occurs immediately following the half second silence in M's utterance. Clearly, this silence was misinterpreted as a turn transition opportunity by R; the confusion over control of the verbal floor is revealed as both participants begin speaking directly following the silence.

4.3.1.2 Question-answer sequences

Another circumstance under which a listener may conclude that a speaker's turn has ended is when the speaker poses a question; the speaker's turn at talk is normally presumed (Sacks, Schegloff et al., 1974) to end immediately after posing a question. Verbal turntaking breakdown occurs when this regularity is violated, as illustrated in the following segments:

•	M: ohh[1] nonono: lets see (.9) wh*eres point bee* R: right there ((2)) M: right [3]her::e? -theres nothing connected to[4] it [R: *yeah::* (.7) R: hmm	1- M jerks cursor to V1 2- M snaps to LB and stares (.9) 3- R snaps to LB 4- R snaps back to WS	

Segment: FF5p13

•	((1)) R: *blood flow* [2]and pressure? (.7) like thats *what this* [3]*whole excercise is about, right?* M: *okay:::- but the::- its says theres two --=* [R: so+ --- where are we sp+ M: =*kinds of gauges:: says* [4]*attach a pressure gauge to the heart at point ay- .hhh- and flow:: gauges at point* bee and cee (.9)	1- M raises to WS (1.2) 2- M drops to LB 3- R rolls cursor randomly around construction, then scribbles it around in WS as he talks 4- R drops to LB 5- R rolls cursor to V	

Segment: AV3p26

The Verbal turntaking breakdown in segment FF5p13 occurs immediately following M's question "right here?", as M's continuing utterance is overlapped by R's answer to the question. The exchange in segment AV3p26 illustrates how a similar breakdown can occur in the latter half of a question-answer sequence, when the expectation that control of the verbal floor should return to the original speaker after an answer is given is violated.

4.3.1.3 Turn ending markers

Another pattern of behavior that regularly implies (Beach, 1993) a turn transition opportunity is when a speaker ends an utterance with a marker like “okay” or “so.” An example of Verbal turntaking breakdown resulting from the misinterpretation of such an utterance is presented in the following segment:

•	(.8) R: as pressure *in the heart* increases-okay (well[1] - then it s gonna increase: then::) [M: okay we [2]have to watch the arrows -- we= M:=have to watch the arrows though [3]<u>watch</u> [R: (though where)	1- uses pen to point to H1, then G1 2- M jabs at H1 3- M is pointing and holding below H1	

Segment: FF2p35

In sum, three patterns of behavior were recognized as turn transition opportunities in this analysis: silences, question-answer sequences, and the use of turn ending markers. Episodes of overlapping talk that occurred immediately following these turn transition opportunities were presumed to reflect confusion over control of the verbal floor, and were consequently identified as Verbal turntaking breakdowns.

4.3.2 Distinguishing Breakdown from Willful Interruption

One issue that remains to be addressed in order to operationalize the criterion for recognizing Verbal turntaking breakdown presented earlier is how soon after a turn transition opportunity overlapping talk must occur in order to evidence turntaking confusion. As discussed earlier, the premise underlying this criterion is that confusion over whose turn it is to speak arises from the misinterpretation of a turn transition opportunity, leading to a situation in which both participants believe they have control of the verbal floor; this confusion becomes evident as both participants begin speaking after the turn transition opportunity. In most cases, however, participants do not begin speaking *precisely* at the same moment; rather, the overlapping utterances are skewed somewhat, with one speaker starting his or her utterance before the other. This raises the question of how long this delay can be before we can assume that the overlapping speaker is fully aware that the original speaker controls the verbal floor, and is *willfully interrupting* that speaker's turn at talk. In this case, the overlapping talk clearly does not evidence underlying confusion over whose turn it is to speak. Rather, the interruption is used by the overlapping speaker as a conversational mechanism (Sacks, 1992a; Sacks, Schegloff et al., 1974; Schegloff, 1987) either to request control over the verbal floor, or to somehow provide feedback to the speaker about the utterance currently being produced. For example, consider the following exchanges:

•	M: *point bee* [5]*is*:: ri::ght [6]her:::e
	((7))
	R: well::: I wouldn't say that there (<u>would</u> be)
	                                    [
	M: wait [8]- there:: is a flow
	    rig*ht* <u>*there*</u>
	((9))
		5- R snaps to LB
		6- R snaps back to WS, M points to right side of V1
		7- Both stare WS, M holds point then releases as R speaks (1.1)
		8- M jabs his finger at V2, apparently seeing the flow arrows
		9- R drops to LB, M is pointing to LB with pen (.6)	

Segment: FF4p17

●	M: workin no::w ((7)) M: aaah I'm[8] gonna have to move this sucker up [[R: (right) yeah ((9)) R: good move	7- R gazes LB, M rolls cursor to H1 (1.0) 8- R gazes WS, M starts H1 upward about an inch as he speaks 9- M finishes moving H1 (1.0)	9- H1 remains hilighted

Segment: AV2p3

In segment FF4p17, the overlapping utterance produced by M is clearly intended to break R's turn at talk so that M can make her own verbal contribution. Conversely, the two interruptions in segment AV2p3 encourage the speaker to continue, providing supportive commentary on the utterance being produced. Importantly, the overlapping contributions in both of these exchanges clearly represent willful interruptions on the part of the overlapping speaker. This means that they can not be counted as Verbal turntaking breakdowns since there was no confusion over whose turn it was to talk; the overlapping speaker was apparently fully aware that it was not his or her turn to speak at the time the overlapping talk was produced. This raises the vitally important question of how to distinguish overlapping utterances that are instances of willful interruption from overlapping utterances that represent breakdowns in turn management.

The heuristic developed to answer this question in this study focuses on *how soon* after a turn transition opportunity the overlapping speech occurs. In particular, if a speaker's utterance has been in progress for some time since the last turn transition opportunity, then it can be assumed that the overlapping speaker was fully aware that it was not his or her turn to speak, and is producing a willful interruption. This leads to the following criterion for recognizing willful interruptions:

> <u>Criterion:</u> If an overlapping utterance is produced after the a speaker's utterance has been ongoing for more than 4-6 syllables after a preceding turn transition opportunity, then the overlapping utterance must be considered a willful interruption; it does not reflect confusion over whose

turn it is to speak and is therefore not evidence of Verbal turntaking breakdown.

This criterion reflects the assumption that participants are not able to instantaneously perceive that a partner has taken control of the verbal floor and abort their impending utterance. For example, consider the exchange presented in segment AO5p9:

•	R: in it a+ -- at different times? ((2)) M: I think so:: [R: its on the [3]right side (1.2)	2- M finishes writing as lifts to WS as she speaks and grabs mouse (1.2) 3- R makes small gesture to her right with her hand	

Segment: AO5p9

The overlapping utterance produced by R occurs shortly (two syllables) after M begins speaking. Applying the criterion just presented, this overlap does not represent a willful interruption since R did not have timely evidence that M had taken control of the verbal channel after the preceding turn transition opportunity. Consequently, the overlap is taken as evidence that both participants believed to have control of the verbal floor and that, therefore, a Verbal turntaking breakdown has occurred.

4.3.3 Exceptions: Non-Linguistic Verbalizations

Several types of verbalizations were exempted from the criteria laid out in the previous section; overlapping contributions involving these verbalizations were never counted as Verbal turntaking breakdown. Non-linguistic contributions like coughs, clicking of the tongue, and throat clearing were not considered to reflect the belief that the participant producing such verbalizations controlled the verbal floor. Another verbalizations that was accorded special treatment was laughter. Though laughter has been shown to be an important conversational tool (Jefferson, 1984) , it is clear that laughter constitutes a very different kind of verbal contribution than sentential speech. In particular, the significance of laughter arises more from its presence, volume, and extent, rather than from the linguistic interpretation of its content. Accordingly, laughter was considered to be a sort of "background" contribution to the interaction not requiring or demonstrating control of the verbal floor; episodes of overlapping verbal contributions in which one speaker is laughing were never taken as evidence of Verbal turntaking breakdown.

4.3.4 Summary: Verbal Turntaking Breakdown

In order to organize their verbal contributions to a conversation, humans have developed highly-refined turntaking mechanisms for regulating access to the verbal channel. Verbal turntaking breakdowns were defined as failures of this turntaking process, leading to a situation in which participants became confused as to whose turn it was to talk. The evidentiary criteria used to recognize episodes of Verbal turntaking breakdown are summarized as follows:

> Criterion: Verbal turntaking breakdown is evidenced by overlapping talk immediately following a turn transition opportunity. Behaviors recognized as turn transition opportunities were extended silences between utterances, question-answer sequences, and the use of turn-ending markers like "okay" and "so".

> Criterion: Overlapping talk evidences Verbal turntaking breakdown only if the overlapping contribution begins within a space of 4-6 syllables following the start of the overlapped utterance. Overlapping speech that occurs outside of this space is considered willful interruption and not counted as Verbal turntaking breakdown.

4.4 Operationalizing Cursor Turntaking Breakdown

Another basic communicative resource to which participants had to organize mutually exclusive access was the shared cursor in the CVCK workspace. The design of the distributed CVCK environment was such that, if both participants attempted to simultaneously control the cursor, the resulting behavior was inevitably disruptive. Accordingly, participants had to continuously negotiate control over the shared cursor, either tacitly or explicitly passing control of the cursor back and forth between them. Cursor turntaking breakdowns occurred when this turntaking mechanism failed, resulting in confusion over who currently held control of the shared cursor. The two possible modalities of such divergent interpretation are illustrated in Figure 4.5.

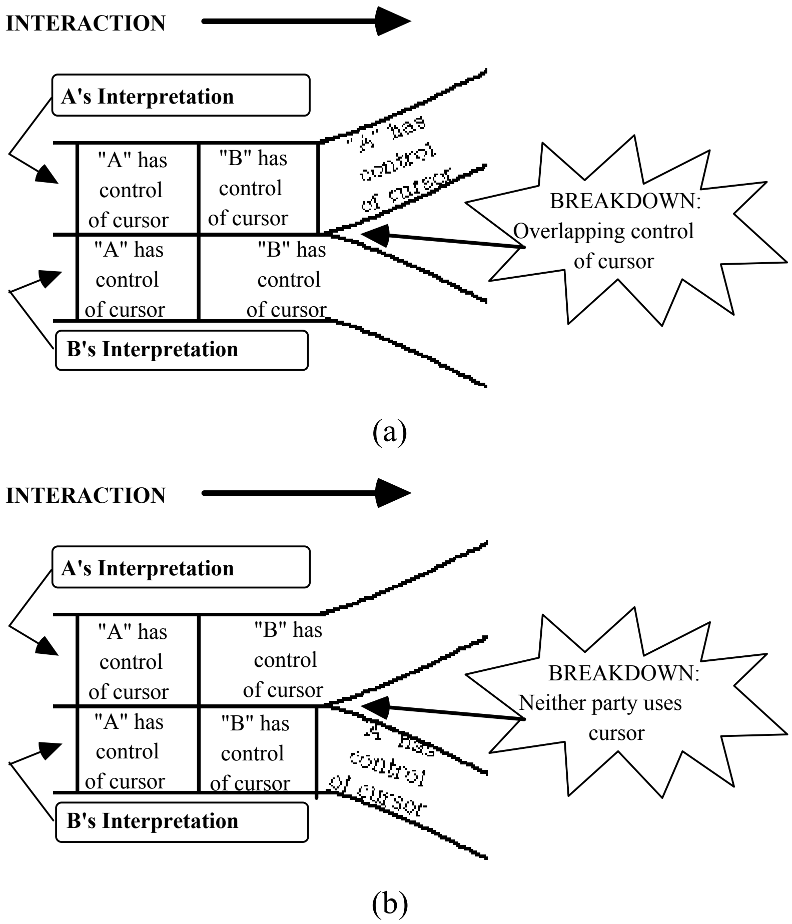


Figure 4.5: Cursor turntaking breakdown is characterized as divergent beliefs over who controls the shared cursor. Either (a) both participants believe they control the cursor or (b) each participant believes it is his or her partner's turn to use the cursor.

Cursor turntaking breakdown has occurred when (Figure 4.5a) there is evidence that both participants have come to simultaneously believe that they have control over the shared cursor, resulting in overlapping attempts to control the cursor, or (Figure 4.5b) when the evidence implies that each participant has come to believe it is his or her partner's turn to use the shared cursor, resulting in an extended period of cursor inactivity. The following sections discuss the evidentiary criteria used to recognize each of these modalities of Cursor turntaking breakdown.

4.4.1 Cursor Turntaking Breakdowns Evidenced by Overlapping Control

The most prevalent form of Cursor turntaking breakdown is when both participants apparently believe they control the cursor, resulting in simultaneous attempts to use the shared cursor to accomplish some action within the shared workspace. Due to the design of the workstation used to implement the shared CVCK environment, these episodes of overlapping control resulted in certain erratic behaviors of the shared cursor and the CVCK environment that were readily apparent in the videotape record. In order to understand the nature of such behaviors, it is necessary to take a moment to explain the way in which the CVCK environment processed the input of the two mice used by interacting participants to control the shared cursor.

Like most personal computers, the Macintosh used for this project is not designed to support multiple independent cursors. Though it is possible to connect two mice to the input bus[8], both mice control a single cursor on the workstation screen. In particular, the device driver associated with each mouse posts "events" to the system's main event queue, denoting changes in the mouse's displacement and the state of the mouse button, as illustrated in Figure 4.6.

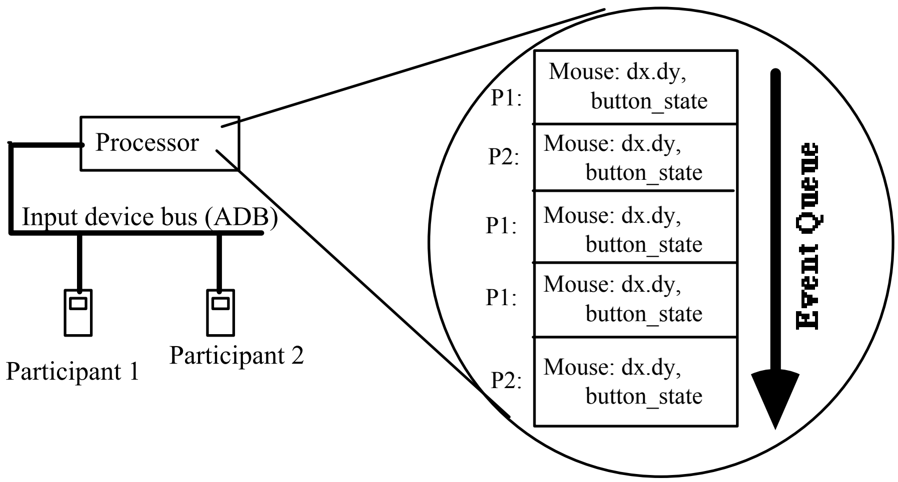


Figure 4.6: How overlapping control of the shared cursor results in easily-observable erratic system behavior. The labels (e.g. P1, P2., etc.) to the left of the queue indicate which participant's mouse posted the event.

When both participants simultaneously move their mouse or click their mouse buttons, the events posted by each participants mouse are interleaved in the event queue. As the system processes these events to update the position of the shared cursor on the screen, the cursor jerks erratically as the system combines the two different displacement vectors posted by the two mice[9]. Even more unusual behaviors result if one participant is holding down the mouse button during a period of overlapping control. In this case, the state of the mouse button (e.g. up or down) fluctuates rapidly as events from the two mice are alternately processed, causing the system to perceive a prolonged series of mouse clicks. The effect of such clicks on the CVCK simulator varies depending on the position of the cursor at the time the phenomenon occurs, but can range from the unexpected creation of new components, to the opening of dialog boxes, to the accidental grabbing and dragging of components within the CVCK workspace.

In sum, the design of the hardware guarantees that overlapping attempts to control the shared cursor will be readily apparent to the analyst. This observation leads to a very straightforward criterion for recognizing Cursor turntaking breakdowns:

> Criterion: Cursor turntaking breakdowns during which each participant believes to have control of the cursor are evidenced by erratic behavior of the shared cursor or simulator resulting from simultaneous attempts by participants to use their respective mice.

An important feature of this criterion is that *all* instances of overlapping control were considered to be evidence of Cursor turntaking breakdown. In particular, no allowance was made for willful interruption of a participant's turn at cursor control. Two observations justify this decision:

1. Overlapping cursor control is inevitably disruptive. Unlike verbal interruptions, which rarely destroy the sensibility of the current speaker's talk, the erratic behavior of the shared cursor when simultaneously controlled by both participants prevents either participant from accomplishing anything constructive. In short, there is no communicative advantage to usurping control of the cursor.

2. Participants have much more effective ways to request control of the cursor. Unlike verbal turntaking, where verbal interruption is the *only* direct way to break into a partner's turn at talk, participants can verbally interrupt a partner's turn at the shared cursor to request control.

Thus, it was assumed that all instances of overlapping control represent confusion over whose turn it currently is to control the cursor and, therefore, can be counted as Cursor turntaking breakdowns.

To illustrate this discussion, the following segments present several examples of Cursor turntaking breakdown:

   •	(.9) R: *pr*essure gauge ---- so a:: pressure ga[5]uge (.9) M: wouldn't---[6]what must be what we= [R: is one of these M: = used before (.4) R: yeah its one of those M: tho::se, [7]why R: *(umlumm)* M: thats a <u>flo</u>w [8]one (.8) if anything=	5- R rolls cursor to V 6- cursor jumps around as both control, settles more or less on V 7- R drops to LB 8- R raises to WS	

Segment: AV3p26

 • •	((1)) M: *that*s like[2] on the dow:::n--- o::h no:: -- SARAH TE+! [[R: the oh AAAAH! ((3)) M: WHATCHA DOING?! -- uhhhhu-h-h [R: I DUNNO! M: .hhh- how can you erase it -uhhhu-h R: but[4] wait -- we can er+ - I didn't do that ((5))	1- R is holding cursor on L staring WS, M gazes LB (.5) 2- M starts clicking and dragging 3- R stops moving mouse, M rolls to WS and clicks near the La and Lb (.6) 4- Both are moving cursor again, so it jerks around, M finally moves it over to C and clicks then to WS and clicks 5- M takes cursor and drags Lb (.7)	2- CVCK perceives multiple clicks. Another L appears partially overlapping first L. Call them La and Lb 3- Lb unhighlights .

Segment: AO3p8

In segment AV3p26, the Cursor turntaking breakdown occurs as both participants speculate which of the icons in the workspace might be the "gauge" referred to in the laboratory manual. A more troublesome exchange involving two Cursor turntaking breakdowns is shown in segment AO3p8. In both cases, participants' attempts to simultaneously control the shared cursor are clearly evident in the erratic behavior of the cursor, and are plainly noted in the transcript of non-verbal behavior.

4.4.1.1 Exceptions: Overlapping Control that does not Evidence Breakdown

The criterion presented in the previous section is based on the premise that all instances of overlapping cursor control represent evidence of confusion over which participant was in control of the cursor and should therefore be recognized as Cursor turntaking breakdown. Two exceptions to this heuristic were made, however, to exempt two relatively uncommon situations in which it was overwhelmingly obvious that an episode of overlapping control was *not* the result such confusion. First, clearly unintentional movements of the mouse (e.g. bumping it with an elbow while turning the page of the laboratory manual) while a partner is controlling the shared cursor were not counted as Cursor turntaking breakdown. Second, episodes like the one presented in segment FF5p1, in which participants explicitly conspire to simultaneously control the cursor were exempted as well:

•	M: I guess ---uh *lets-- see if you can use* [1]*your mouse at the* [2]same ti::me (.6) R: ahh -- ohh ((3)) R: no:: M: oh wait we <u>fi</u>ght eachother I gue::ss ((4)) R: are we:: sposed to fight eachother? ---- no we're --- (we+) -- we haffta to **coopera**te [M: (less see) ---you try it M: *you try it* now::	1- M is actually gazing Rs mouse right next to LB, here M makes a small finger gesture pointing at mouse; R gazes WS just in time catch it 2- R gazes WS grabbing for mouse 3- Cursor jumps erratically as both move mouse (1.1) 4- R stops controlling mouse and M rolls cursor to palette (.9)	6- New V appears, highlighted

Segment: FF5p1

4.4.2 Cursor Turntaking Breakdowns Evidenced by Explicit Verbal Repair

As indicated in Figure 4.5b, another modality of Cursor turntaking breakdown occurs when both participants come to believe that their partner currently controls the cursor, resulting in a period of extended inactivity as each of them waits for the other to act. In general, it impossible to reliably recognize this condition, since it requires distinguishing periods of inactivity due to Cursor turntaking breakdown from mundane pauses in cursor activity that occur throughout an interaction. The only situation in which this form of Cursor turntaking breakdown becomes unambiguously evident in the record of interaction is when the underlying confusion over cursor control is explicitly recognized and repaired by participants. This leads to the following criterion for recognizing Cursor turntaking breakdown:

> Criterion: Cursor turntaking breakdowns in which each participant believes it is the other's turn to use the shared cursor are evidenced by an explicit verbal repair immediately following an extended silence, in which the issue of who should control the cursor is explicitly raised and resolved.

The following segments present several examples of Cursor turntaking breakdowns identified by this criterion:

●	M: ok-run it again[3] and lets [4]keep a good eye on the valves= R: =lets do-more beats than one M: okay (1.0) R: do you wanna change that or do you want me to. M: go ahead ((5))	3- R raises to gaze WS 4- M takes hand off of mouse while R soon (.5) reaches for hers. 5- R clicks on the beats box, then go over to the slider and clicks on that (4.8)	5- when clicked, beats box hilights, but unhilights when slider is clicked. slider slides a bit

Segment: AO2p8

	R: *mmmmmmmmmm ---- it says modify it* *and then in - number two it says =* [M: *o::kay* R: =*convert the gauges t+--- into graphs -* so --- double click on em ((4))	4- M raises to gaze WS and both stare idly (2.2)	

(continued)

●	R: go for it ((1)) M: oh ((2))	1- M sits up and grabs mouse (.5) 2- M rolls cursor down onto G1 (1.1)	

Segment: AO5p16

●	((4)) R: maybe it doesn't matter (.7) M: hmmm okay:::- try that[5] R: okay:: (.5) go ahead M: *unkay - I'll do* [6]bee:: R: *okay* ((7))	4- R stares WS, M drops to gaze LB, then back to WS (2.6) 5- M drops to LB 6- R drops to LB as M grabs cursor and recenters on gauge icon 7- R raises to WS (.6)	

Segment: AO4p19

In each of the above segments, participants' agreement on some course of action is followed by an extended silence, during which each participant believes that his or her partner has control of the shared cursor and will perform the action agreed upon. The Cursor turntaking breakdowns are revealed when the confusion over who controls the cursor is explicitly repaired through a verbal clarification of who should take control of the cursor and perform the action.

4.4.3 Summary: Cursor Turntaking Breakdown

Manipulating the CVCK simulator using the shared cursor was the central activity in the task assigned to participants in the interactions analyzed in this study. Because simultaneous attempts to control the shared cursor resulted in chaotic behavior, participants were strongly motivated to negotiate mutually exclusive access to the shared cursor, tacitly or explicitly passing control over the shared cursor back and forth between them. Cursor turntaking breakdowns occurred when this turntaking process failed, leading to divergent beliefs over which participant was currently in control of the shared cursor. The following evidentiary criteria were established to identify episodes of Cursor turntaking breakdown:

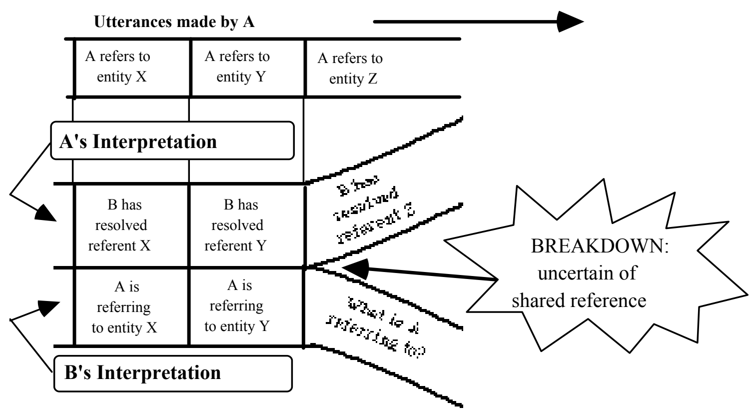
> Criterion: Cursor turntaking breakdowns in which both participants believe they currently control the cursor are evidenced by erratic behavior of the shared cursor or simulator resulting from simultaneous attempts by participants to use their respective mice.

> Criterion: Cursor turntaking breakdowns in which each participant believes it is the other's turn to use the shared cursor are evidenced by an explicit verbal repair immediately following an extended silence, in which the issue of who should control the cursor is explicitly raised and resolved.

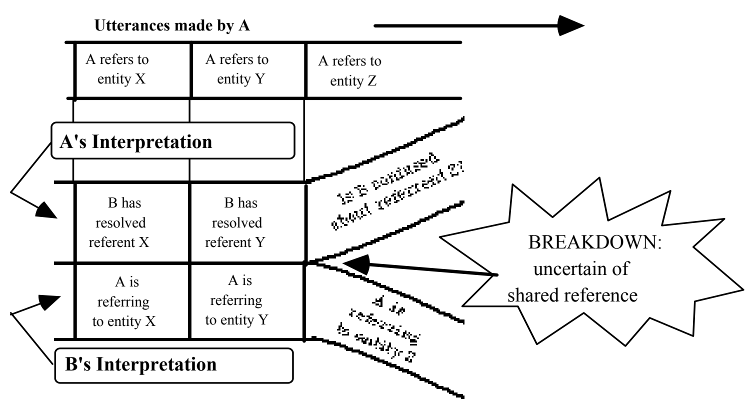
In sum, nearly all instances of overlapping cursor control were taken to stem from underlying confusion over which participant currently controls the cursor, and were therefore counted as Cursor turntaking breakdowns. The only exceptions to this heuristic were instances in which the overlapping cursor control was due to accidental contact with the mouse, or when participants explicitly conspired to simultaneously control the shared cursor.

4.5 Operationalizing Reference Breakdown

As participants in any interaction converse, they must continually track the indexical significance of each other's utterances, accurately matching the references that appear in each utterance with entities that exist in the referential context. For example, in the utterance "Let's move the pump thing over to the side," there are at least two references which must be disambiguated in order to construct the meaning on the utterance: which object is the referent of "the pump," and what spatial position is meant by "the side." In negotiating shared reference, the listener examines the current context, searching for the appropriate referents, while the speaker monitors the listener's verbal and non-verbal displays for evidence that the listener has correctly resolved the references. Participants have been shown to rely on a wide variety of conversational regularities and resources (Anderson, Bader et al., 1991; Clark & Schaefer, 1989; Clark & Wilkes-Gibbs, 1986; Heath, 1986; Isaacs & Clark, 1987; Wilkes-Gibbs & Clark, 1992) to establish and maintain shared reference as each new utterance is produced over the course of an interaction. Reference breakdown is defined by the failure of this negotiated process, resulting in evidence that there is some uncertainy over whether a reference in a speaker's immediately preceding utterance was accurately resolved by both participants.



(a)



(b)

Figure 4.7: Reference breakdown is characterized by confusion in either the speaker or the listener about whether shared reference has been established. Assume that "A" is the speaker.

As illustrated in Figure 4.7, there are essentially two ways in which such confusion can arise.

The most obvious form of Reference breakdown (Figure 4.7a) is when a speaker makes reference to some entity in the shared context and the listener implies that he or she has been unable to unambiguously locate the referent by bringing the uncertainty to the attention of the speaker and soliciting further help in locating the referent. A second form of Reference breakdown (Figure 4.7b) occurs when the speaker displays uncertainty over whether an immediately preceding reference was understood by the listener by spontaneously producing a repair utterance to somehow refine the original reference.

These two scenarios of referential confusion — requested repair and self repair — define the interpretive framework used to recognize Reference breakdowns in this analysis. The following sections describe the evidentiary criteria developed to identify each of these forms of Reference breakdown.

4.5.1 Reference Breakdown Evidenced by Requested Repair

One way in which Reference breakdown is evidenced in the observable record of interaction is when a listener initiates a repair sequence in which the listener's uncertainty about the referent of the speaker's immediately preceding utterance is made available to the speaker and collaboratively resolved. This leads to the following criterion for identifying Reference breakdown:

> Criterion: Reference breakdown is evidenced by an explicit verbal repair sequence initiated by a listener, aimed at clarifying the referent of an utterance just produced by the speaker.

As illustrated in the following segments, the repair work required to resolve referential ambiguity is often minimal, consisting of a simple question-answer sequence in which the listener presents a possible referent to the speaker for confirmation:

	((5)) M: What [6]about the[7] second thing down	6- M finger points to WS 7- R gazes WS and grabs mouse	
●	R: (um that) [1]this one? M: yeah:: ((2))	1- R rolls cursor down to V in pallete 2- R does a “describe” on V, they read it (5.0) then R while gazing LB (1.0)	2- Describe dialog for V pops up. V stays hilighted.

Segment: AV5p3

●	M: okay:: then throw one[6] right the[7]:::::re ((8)) R: where (.7) M: on the[9]::: second one ---yea::h *right there* [R: this one?	6- R rolls up and drags in another gauge 7- M points with pen at V1 8- R drags G down below V1 (.7) 9- M points more specifically to the right side of V1 while R fluidly slides the G over to the second rightmost port on V1 10- M raises to WS again as R carefully positions G (2.8)	6- new G appears in workspace

Segment: FF4p8

	M: yeah::: (.6) so *I I would say those are open* [R: whats [7]the step thing	7- R rolls cursor to STEP	

•	(.5) M: *huh::?* R: whats --[1]-- step (.7) M: *step?* R: up here[2] at the top ((3)) M: ohhhh, at the top? (.4) R: yeah::, see? [M: ohhh ((4)) R: where the cursor is? M: mhmm (.5) R: you know what that means?	1- M hunches forward and examines LB intently 2- R is wiggling cursor over STEP 3- M gazes LB, then raises to WS to speak (1.6) 4- R wiggles cursor again, M appears to see it already (.8)	

Segment: AO4p12

In segment AV5p3 and FF4p8, the Reference breakdown is revealed as the listener displays uncertainty over an immediately preceding reference by deictically identifying a possible referent and explicitly requesting confirmation from the speaker; the repair is trivially accomplished as the original speaker verifies that the listener has, indeed, located the correct referent. Segment AO4p8 presents an episode of Reference breakdown which required a more lengthy repair to clarify the R’s reference to “the step thing”; shared reference is eventually established after M realizes that R is talking about something in the shared workspace, rather than in the laboratory manual, and perceived R’s deictic gestures with the shared cursor.

A final example of Reference breakdown evidenced by requested repair is presented in the following segment, in which the repair sequence initiated to resolve a referential ambiguity is itself hampered by a further Reference breakdown:

•	M: [5]look at that one -- dyou see how it -- {didlululun} --- see how it tingles a little? -[5]-- it goes {didilidlidle} R: the one:: by the ----- the one where the, = [M: (duli+) R: = things crooked?	5- M clicks RUN	5- CVCK runs another cycle
•	((1)) M: which[2] things crooked-uh<u>uh</u>-h (.5) R: wher+ -- ohh the little (efffuh) [M: the <u>bi::g</u> - like okay:: the-*the <u>top</u> of it?* R: [3]yea::h M: that big b::::::: ---- blubble thing and then right to the right of that R: okay	1- Both stare WS (1.9) 2- R clicks RUN 3- M snaps back to WS	2- CVCK runs a cycle

(continued)

	((4)) R: (e+) [M: like right here ------5 right there R: it ---- shake6s M: yea::h ((7))	4- M rolls cursor from RUN down to V2 (.5) 5- M centers cursor on V2 6- R rolls cursor back up towards control panel 7- Both watch WS as R clicks RUN (1.9)	7- CVCK runs a cycle

Segment: AO3p19

In segment AO3p19, an initial Reference breakdown is revealed as R initiates a repair sequence in which she verbally describes a possible referent of M's immediately preceding utterance (i.e. "the one where the things crooked?"). A second Reference breakdown is evidenced as M displays uncertainty over which entity is being referred to as the "crooked thing." Both confusions are eventually repaired when M refines her original reference, reinforcing her revised reference with deictic gesture using the shared cursor.

4.5.2 Reference Breakdown Evidenced by Self Repair

As illustrated in Figure 4.7b, another form of Reference breakdown occurs when a speaker displays uncertainty that the listener has correctly interpreted a reference made in an immediately preceding utterance by spontaneously providing a repair aimed at clarifying the referent. This self repair behavior provides the evidentiary basis for the following criterion for recognizing Reference breakdown:

> Criterion: Reference breakdown is evidenced by a spontaneous repair provided by the speaker of an immediately preceding utterance, aimed at further clarifying the referent of that utterance.

The following segments present several examples of Reference breakdown evidenced by self repair:

	((7)) R: whatter [8]the::se	7- M clicks in STOP then STEP (1.3) 8- R points with finger at H1, then drops down to tap on C1	7- CVCK runs one step
•	((1)) R: whatter uuuuh [2]--- whats - whats this thing here dyou spose[3] - and this thing ((4))	1- R finishes pointing and turns to RS, then back to WS and grabbing mouse as he speaks; meanwhile M gives cursor a final jerk and stares WS (2.0) 2- R rolls cursor over to H1 3- R rolls cursor to C1 4- M drops to LB (.6)	

Segment: AV2p11

•	M: wait -- go for this one[1] right ther::e R: [2] {oughh} ((a grunt)) M: oh+ ----- that[3] one (.9) number four down ((4)) M: achh! ---- dont move that! --- it goes it the middle.	1- M rolls cursor down to U, then it jerks as both control it 2- cursor circles wildly as R “scribbles” mouse 3- R rolls cursor to U and then drags out new U during subsequent pause 4- R positions new U near H1, M glances LB and back; M apparently taps mouse as he passes over H1, causing bobble in mousedown condition (2.3)	3- U highlights, then new U appears in WS 4- U is dropped and unhighlighte d, H1 highlights and is dragged

Segment: AV3p2

•	M: thats true R: uu::::::mm M: isn’t it? [R: I don’t--know cuz [3]look at the graph[4]-- the second gra:ph[5]? (1.4) M: unkay R: theres spots--[6]--where theres no flow	3- R raises hand/pen to point to the g2 in WS 4- M drops quickly to LB 5- Now M gazes WS again and waits for R to continue 6- R taps the screen at g2 to point out the spots.	

Segment: AO2p17

In segment AV2p11, R's uncertainty over his previous reference to "these" is evidenced in the repair subsequently produced, in which R rephrases his earlier reference, adding deictic gesture with the shared cursor to identify the referents of the utterance. A similar spontaneous repair (i.e. "number four down") is produced by M in segment AV3p2 after a persistent Cursor turntaking breakdown prevents her from deictically identifying the referent of her original utterance. Finally, in segment AO2p17, M's uncertainty over the adequacy of her reference to "the graph" becomes apparent as she spontaneously revises her reference to specify "the second graph".

4.5.3 Domain-Related Referential Ambiguities

In focusing on the communicative interaction of the two human participants, it is easy to forget that there are actually two other "participants" in the interaction, at least when it comes to making references: the CVCK system and the laboratory manual. Both of these artifacts contain or produce textual instructions or descriptions that refer to entities in the participants' communicative context. Because the focus of this analysis was on how well the participants were able to maintain shared interpretations of the referential significance of *each other's* communicative displays, confusions experienced by participants over references made by either of these artifacts were not counted as Reference breakdowns. The following example illustrates the kind of confusions ruled out by this restriction:

●	M: *okay:::* (.6) *there are times when there is flow:: toward the heart* at point bee[2]:: ((3)) R: wheres the heart (4.0) M: uuumm ((4)) M: I th<u>ink</u>[5] <u>tha</u>ts sposed to be the heart (3.8) R: they dont make that clear: (.7) but yet	2- R raises to WS -- nearly in synch 3- Both just stare at WS (1.8) 4- R drops to LB, then snaps to WS as M leans to point just before speaking (1.6) 5- M jabs pen at H1 and retracts	

Segment: FF5p13

The referential confusion in segment FF5p13 occurs as a result of the reference to the "heart" made in the laboratory manual instruction read by M in the first utterance; both participants are confused about which of the icons in the shared workspace is referred to by the instruction. Clearly, this referential ambiguity stems from the minimalist design of the laboratory manual and has nothing to do with the efficacy of communicative interaction between participants.

4.5.4 Summary: Reference Breakdown

The sensibility of any conversation depends on the ability of participants to maintain shared reference, continuously constructing the referential relationship between linguistic references and entities in the shared referential context. Reference breakdown occurs when either one or both participants become uncertain that shared reference has been established, resulting in an explicit verbal repair aimed at resolving the ambiguity. Such repair may be initiated by either the speaker or the listener, leading to the following evidentiary criteria for recognizing Reference breakdown:

> <u>Criterion:</u> Reference breakdown is evidenced by an explicit verbal repair sequence initiated by a listener, aimed at clarifying the referent of an utterance just produced by the speaker.

Criterion: Reference breakdown is evidenced by a spontaneous repair provided by the speaker of an immediately preceding utterance, aimed at further clarifying the referent of that utterance.

Both of these criteria are quite conservative, requiring referential confusions to be revealed through explicit verbal repair sequences in order to be recognized as Reference breakdown.

4.6 Operationalizing Topic Breakdown

The notion of "current topic" is a vital interpretive resource for constructing the sensibility of ongoing action. A participant's conception of what the current topic of discussion is fundamentally shapes the interpretation of the communicative displays of collaborating partners, as well as informing expectations about upcoming action. As an interaction progresses, the current topic of conversation is continuously negotiated by participants as new topics are introduced and oriented to by participants, and discussion progresses fluidly from topic to topic.

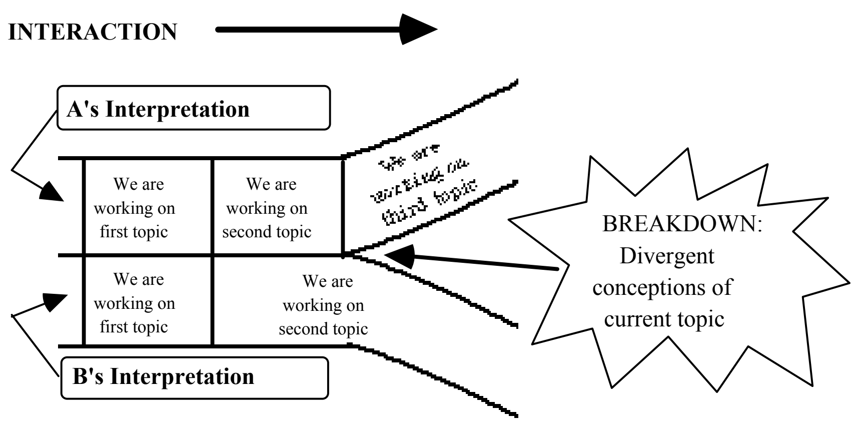


Figure 4.8: Topic breakdown characterized as divergent interpretations of current topic of collaborative discussion.

Topic breakdown occurs when there is evidence that participants' conception of "current topic" has diverged, leading to an asynchrony in topical orientation in which one participant believes that discussion has moved to some new topic, while the other believes that discussion remains focused on the previous topic, as illustrated in Figure 4.8.

In order to understand how Topic breakdowns were evidenced in the interactions analyzed in this study, it is necessary to briefly discuss how topic management in task-oriented interactions differs from that in mundane personal conversations.

Most studies of how participants negotiate topic over the course of an interaction (Beach, 1990; Beach, 1993; Button & Casey, 1984; Covelli & Murray, 1980; Jefferson, 1993) have focused on mundane personal interaction in which participants are engaged in conversations whose goals and extent are defined primarily by the participants themselves. In these scenarios, new topics are verbally introduced by one participant and subsequently oriented to by other participants. Though individual participants may have certain "agendas" of topics they wish to raise, they are generally introduced opportunistically rather than serving as a formal framework for the discussion. The fact that each new topic must be verbally introduced by one of the participants makes Topic breakdown relatively unusual in personal interactions; the only way that participants' interpretation of topic can diverge is if a participant somehow fails to recognize a partner's utterance as a new topic.

By contrast, the way in which participants manage topic in task-oriented interactions has received limited attention until quite recently (Fox, 1993; Whalen, 1995) . In task-oriented interactions, the overall topical structure of the interaction as well as its goals and extent are defined by the task solution process itself; accomplishing the task requires performing a predefined series of activities that constitute the main topics of the interactions. Of course, participants may also spontaneously introduce unique new subtopics of discussion, but all such topics exist within and are understood in reference to the overall topical structure defined by the task solution process (Fox, 1993) . In terms of topic management, the most important consequence of having an overall topical structure that is known to both participants is that, at any given point in the discussion, the "next" topic of discussion is predefined by this structure. In particular, the next topic does not necessarily have to be verbally introduced; the interaction can tacitly move to the next topic "by default." This feature of task-oriented interactions makes them much more susceptible to Topic breakdown, in that participants are continually expected to tacitly track the topic of discussion by examining the task-related behaviors of their partner.

The interactions analyzed in this study were clearly task-oriented, with the overall topic structure defined by the laboratory manual[10]. In order to accomplish the task, participants had to sequentially address each of the topics embodied in the instructions given in the laboratory manual, placing various components to build the construction, running the simulator, and answering various questions. Topic breakdowns occurred when one participant's behavior implied that he or she believed the current task or question had been addressed and that the discussion had moved on to the next topic, while the other participant continued to work on the previous topic. The following criterion was defined to identify such divergent interpretations of current topic and operationalize the category of Topic breakdown:

> Criterion: Topic breakdowns are evidenced by explicit evidence in the verbal record of interaction that participants do not have a shared interpretation of what the "current topic" of discussion is. This evidence may take one of two forms: Repair sequences in which the confusion over current topic is explicitly expressed and resolved by participants; and adjacent utterances that clearly reveal that participants are working on different topics.

The following sections discuss each of these two evidentiary patterns of behavior in more detail.

4.6.1 Explicit Topic Repair Sequences

The most obvious way in which confusion about the current topic of discussion is revealed is when a participant initiates an explicit verbal repair in which the issue of “what are we talking about” is somehow raised and collaboratively resolved by participants. The following segments provide several examples of Topic breakdown evidenced by explicit verbal repair sequences:

•	((1)) R: its always[2] flowing towards the heart--it comes out of the heart and goes towards the heart (.4) so its always flowing towards the heheheeart -[3]-- whether your at point C or not M: *this is true* [R: but I know what it means (.8) okay--[4]-the flow graph *for C* ((5)) R: it flows when the pressure is high ((6)) M: *where are you at (.4) your not on part three yet* [7]*(.4) did I miss something?* R: *No- at the bottom of part three* M: *okay*	1- M marks an answer while R looks at LB then WS (3.6) 2- R is gestureing as she speaks, vaguely shaping the in-out flows. 3- M gazes at LB and throws her hands and body back and forth in gestures too. 4- M audibly turns page to next section 5- M finishes turning page, R stares LB, then WS, gestures vaguely at screen with pen (3.9) 6- R drops to LB (1.3) 7- M pages back to look at previous page	

Segment: AO2p21

Segment AO2p21 shows participants just finishing discussion of one topic; a new topic is implicitly introduced as M refers to the next question in the laboratory manual (i.e. “okay -- the flow graph for C”). As it turns out, however, M has accidentally skipped the last question on the current page and has moved on to the next page of the laboratory manual. This divergence in topical orientation is revealed when M explicitly raises the issue of current topic by asking “where are you at,” prompting R to clarify her current topical orientation.

•	((4)) M: *oh it has to have a valve closed to build up pressure* ((5)) M: *right?* ((6)) M: what are you doing ((7)) R: I'*m just playing*	4- M marks LB, glancing up once, as R gazes WS and clicks RUN 6 times slowly (6.8) 5- M marks, R still clicking RUN about 5 more times (2.2) 6- Same as 5, about 3 more RUNs, then M pauses and raises to LB to speak (1.6) 7- M drops to LB and continues marking, R clicks RUN several more times, then drops mouse, grabs pen (2.0)	4- CVCK runs six cycles in a row 5- CVCK runs a cycle for each click 6- see 5 7- see 5

Segment: AV3p18

In segment AV3p18, M is apparently unable to discern the relevance of R's actions using the shared cursor with respect to what she believes to be the current topic of their collaborative interaction, causing her to suspect that a R is working on some topic unknown to her, and prompting her to initiate a repair by asking "what are you doing?"

The exchanges presented in the two preceding segments are both clearly identified as Topic breakdowns by the criterion presented earlier — the underlying confusion over current topic is unambiguously evidenced in the verbal repair sequence initiated by a participant.

4.6.2 Trivial Topic Repair

In the exchanges presented in the preceding section, participants' perceptions of current topic were resynchronized via an explicit repair sequence in which the issue of "what topic are we on" was explicitly introduced as a digressionary topic in the discussion. In many situations in which a Topic breakdown has occurred, however, no such extensive effort is required to resynchronize participants' conceptions of current topic. Specifically, participants' differing topical orientation may be implicitly evident in the content of their utterances. For example, consider the following exchange:

•	M: okay:: so now it go:::es - clockwise[3] R: (u) ---- (es) ((4)) M: *uuhho+ gaw::::d* ((5)) M: *okay -- .hhh- when blood* [R: no it [6]doesn't (.5) M: *WH*AT[6]?! R: no[6] it doesn't M: it doesn't go clockwise	3- M drops to LB and marks, R gazes WS 4- R starts to LB, but aborts and grabs mouse; M is audibly scribbling out something in LB; R clicks RUN just as M speaks (2.3) 5- R clicks RUN and watches two more times as M marks, then drops pen and speaks (2.4) 6- R clicks RUN	4- CVCK runs a cycle 5- CVCK runs two cycles 6- CVCK runs a cycle

Segment: AV3p15

This exchange begins with M proposing an answer (i.e. "so now it goes clockwise") to the question that participants are currently working to answer, which asks participants to determine the direction of blood flow in the cardiovascular loop. Subsequently, M turns to mark the answer in the laboratory manual, apparently considering the topic closed, as R continues to work on the question by running the simulator several more times. The Topic breakdown is unambiguously evidenced when R continues discussion of the previous topic by asserting that blood flow is not clockwise, while at the same time M begins to read the next question from the laboratory manual. However, no explicit repair sequence is ever initiated; based on R's utterance, M is able to infer that R is still on the previous topic, trivially repairing the difference in topical orientation.

4.6.3 Summary: Topic Breakdown

The concept of topic constitutes a primary organizational mechanism for structuring interaction, establishing an interpretive framework for constructing the significance of action and defining the basis for understanding what it means for interaction to "progress." As an interaction evolves, participants must continually work to maintain a shared topical orientation as new topics are introduced, addressed, and closed. Topic breakdowns occur when participants develop differing beliefs regarding the current topic of their interaction. Such divergences in topical orientation were recognized using the following evidentiary criterion:

> Criterion: Topic breakdowns are evidenced by explicit evidence in the verbal record of interaction that participants do not have a shared interpretation of what the "current topic" of discussion is. This evidence may take one of two forms: Repair sequences in which the confusion over current topic is explicitly expressed and resolved by participants; and adjacent utterances that clearly reveal that participants are working on different topics.

By insisting on verbal evidence of topical confusion, this criterion emphasizes that non-verbal evidence of topical confusion alone (e.g. turning to the next page in the laboratory manual) does not unambiguously evidence Topic breakdown, since there is no way to be certain that participants' conception has actually diverged. At the same time, the insistence on verbal evidence of topical confusion does not necessarily imply an explicit repair sequence in which the topical confusion is collaboratively resolved by both participants; often differences in topical orientation are revealed in the content of adjacent utterances and trivially resolved, as participants realize the asynchrony and both orient to a single[11] topic.

4.7 Summary: Identifying Patterns of Breakdown

The goal of this chapter has been to establish the comparative framework for the analysis of communicative efficacy presented in this dissertation by describing the patterns of communicative breakdown revealed in the initial qualitative phase of this study, and presenting the operational criteria developed to recognize episodes of breakdown in each category. The discussion is graphically summarized in Figure 4.9:

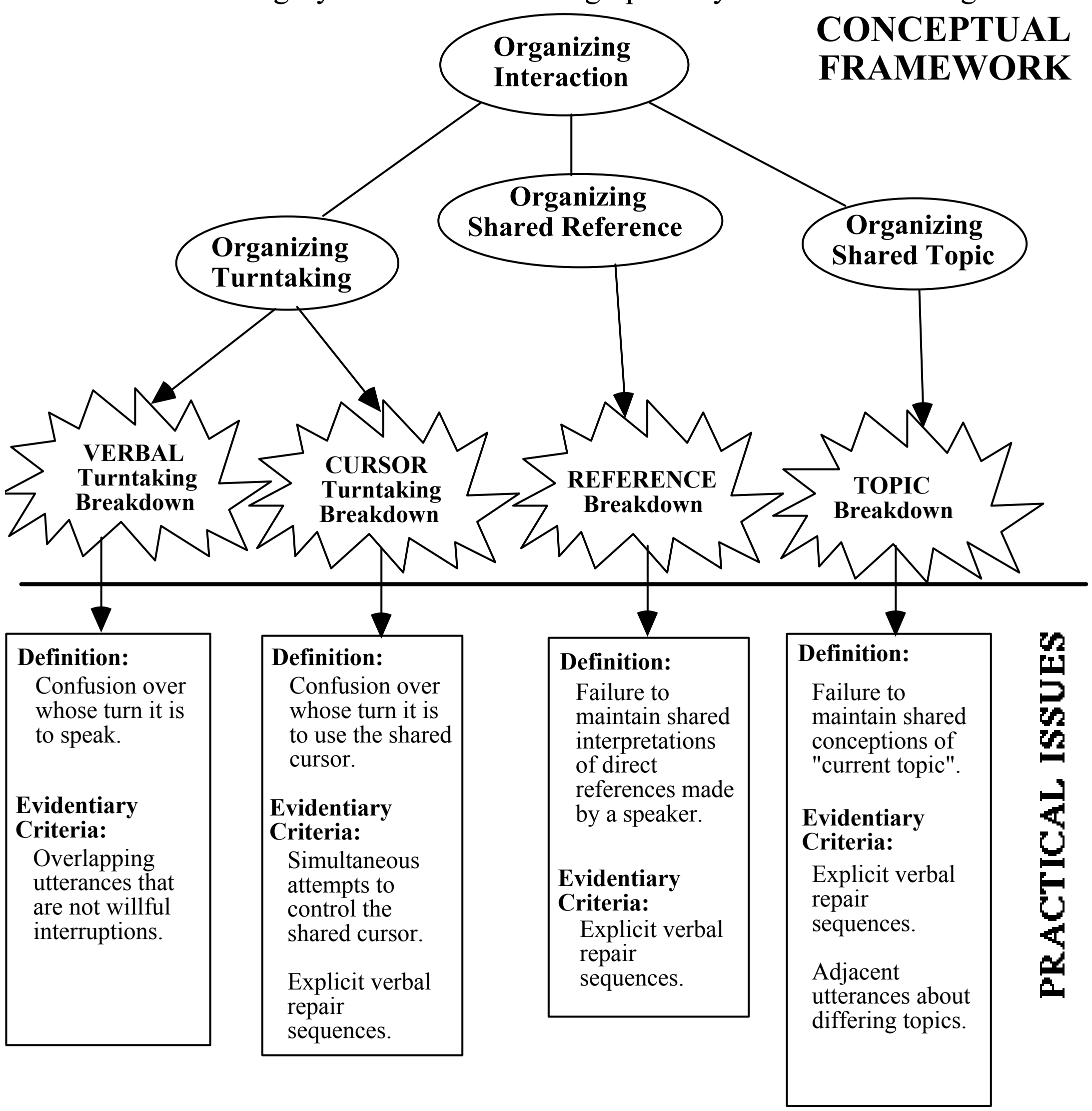


Figure 4.9: Summary of categories of breakdown and criteria for identifying them.

As indicated in Figure 4.9, four specific patterns of communicative breakdown were identified in the analysis: Verbal turntaking breakdown, Cursor turntaking breakdown, Reference breakdown, and Topic breakdown.

Both Verbal and Cursor turntaking breakdown can be seen as resource management difficulties characterized by failures to maintain shared interpretations of which participant's turn it is to make a contribution to the interaction using some mutually exclusive resource: Verbal turntaking breakdowns were related to regulating verbal contributions, while Cursor turntaking breakdowns were associated with regulating access to the shared cursor that participants used to manipulate the electronic workspace. In general, the evidentiary criteria used to recognize instances of Verbal and Cursor turntaking breakdown episodes of breakdown in the record of interaction were primarily focused on locating episodes of overlapping control of the resource in question.

Reference and Topic breakdowns were both related to failures of participants to maintain shared interpretations of the higher-level semantic aspects of their evolving interaction. Reference breakdowns were related to troubles interpreting direct references made by a speaker to objects and entities in the task environment, causing either the speaker or the listener to become uncertain whether shared reference had been successfully established. Episodes of Reference breakdown were evidenced by explicit verbal repair sequences in which referential ambiguity was somehow resolved by participants. Finally, Topic breakdowns were defined as failures to maintain synchronous conceptions of what problem or issue constituted the current topic of interaction, resulting in a situation in which one participant has moved on to the next topic (as defined by the laboratory manual), while his or her partner still believes the previous topic to be the current focus of discussion. Topic breakdowns were detected by the presence of explicit verbal repair sequences in which participants collaboratively resynchronized their topical orientations, and also by adjacent utterances clearly indicating that participants were focused on different topics.

Because recognizing episodes of communicative breakdown is based on a retrospective reconstruction of the significance of participants' behavior from the record of interaction, it inevitably requires some inference on the part of the analyst. In particular, it is impossible to define deterministic heuristics for mechanically identifying instances of breakdown based on the presence of certain context-independent, objectively-defined features of interaction. At the same time, the comparative goals of this analysis make it imperative to establish evidentiary criteria that define the behaviors that represent evidence of communicative breakdown as specifically and consistently as possible. Accordingly, the evidentiary criteria established for operationalizing each pattern of breakdown exposed by this analysis were very conservative, strictly constraining the subjective element by focusing on clearly recognizable behaviors like overlapping control of a resource and explicit verbal repair sequences.

In sum, the four patterns of breakdown and the strong evidentiary criteria developed for recognizing instances of breakdown in each category presented in this chapter establish a strong foundation for assessing the total amount of communicative breakdown that occurred in each of the interactions analyzed in this study, and stochastically comparing the amount of communicative breakdown experienced by participants interacting in different communication environments. It is this quantitative analysis of breakdown that we turn to in the next chapter.

4.8 Notes

[1] The second qualitative phase of Breakdown Analysis, the results of which are presented in Chapter 6, is more typical of interaction analytic studies, working to rationalize differences between environments by examining instances of breakdown in detail.
[2] The framework developed here can be seen as a generalization of the set of analytic foci developed by interaction analysts (Jordan & Henderson, 1995) to shape the analysis of videotaped interactions.
[3] Note that these studies generally focus on the way in which turn management is related to participants' efforts to achieve various higher level conversational goals. For example, Pomerantz (1975) studies the way in which extended silences are used by listeners to *avoid* responding to a speaker's question, implicitly passing control of the verbal channel back to the speaker.
[4] In actuality, a fourth column was used to annotate the transcript with insights and observations of the analyst during transcription. While these annotations were a valuable resource for the analysis, they are not directly related to textually representing action; the fourth column is not shown in any of the segments of transcript presented in this work.
[5] This approach can be seen as an extension to one used in (Schegloff, 1984) .
[6] In the audio-video condition, gaze at the other participant meant gaze directed at the remote video image. Obviously, no gaze at the other participant could occur in the audio-only condition, since no visual connection between participants existed.
[7] This was, indeed, a very common mistake made by participants in the interactions analyzed.
[8] Appendix B provides a detailed review of the technical aspects of the CVCK environment.
[9] Even if participants attempt to move the cursor to the same place at the same moment, the displacement vectors will be slightly different, causing noticeably jerky cursor behavior.
[10] For reference, the laboratory manual given to participants is reproduced in Appendix C.
[11] In nearly all cases, the participant who has moved on to a next topic returns to the previous topic.

CHAPTER V

DIFFERENCES IN COMMUNICATIVE EFFICACY

The central research question addressed in this dissertation is whether communicative interactions in technologically-mediated communication environments are just as effective as interactions in which participants are physically copresent. After developing strong theoretical and practical foundations for this comparative analysis in previous chapters, we are finally prepared to directly answer this question by comparing the total amount of communicative breakdown experienced by participants interacting in the three different communication environments examined in this study. The initial qualitative phase of the Breakdown Analysis, the results of which were presented in Chapter IV, established a framework for comparison by identifying consistent patterns of breakdown that occurred in the interactions examined. Briefly, the four categories of communicative breakdown identified were:

1. <u>Verbal turntaking breakdowns</u> — confusions over which participant currently controlled the verbal floor.

2. <u>Cursor turntaking breakdowns</u> — confusions over whose turn it was to use the shared cursor to manipulate the electronic workspace.

3. <u>Reference breakdowns</u> — failures to maintain shared reference to the objects and entities in the task context.

4. <u>Topic breakdowns</u> — failures to maintain a shared conception of what the "current topic" of discussion was throughout the interaction.

In addition to revealing these four categories of communicative breakdown, the analysis presented in Chapter IV placed a strong emphasis on developing consistent operational criteria for recognizing episodes of breakdown in each of these categories.

This chapter presents the results of the second, quantitative phase of Breakdown Analysis, in which these operational criteria are used to assess the total amount of communicative breakdown that occurred in each of the environments analyzed. Significant differences in the overall communicative efficacy of interaction in the three communication environments are then tested by statistically comparing the number of breakdowns that occurred in the interactions that took place in each environment.

The following section sets the cornerstone for this statistical analysis by briefly reviewing the analytic procedure used to assess the total amount of breakdown in each environment and then presenting the raw results of this enumeration. In section 5.2, nonparametric statistical techniques are applied to test whether significant differences in the amount of breakdown exist between environments. Finally, conclusions about the relative communicative efficacy of interaction in the three environments are drawn based on the results of the statistical evaluation.

5.1 Quantifying Communicative Breakdown

To determine the total amount of communicative breakdown that occurred in each environment, transcripts for all interactions were re-examined in their entirety, applying the criteria developed in Chapter IV to identify all episodes of communicative breakdown in each of the four categories. Several measures were taken to ensure that all episodes of breakdown were identified and that the evidentiary criteria were applied consistently throughout the analysis:

1. Iteration. A total of four passes were made through the data, with each pass confirming previously-identified breakdowns and recognizing further episodes of breakdown that had escaped detection in earlier passes.

2. Interleaving. The order in which interactions were analyzed was varied, alternating between the three communications environments to avoid any possible interpretive bias resulting from sequentially examining several interactions that took place in the same environment. For example, analysis of a copresent interaction was always followed by analysis of an audio-only or audio-video interaction; analysis of an audio-only interaction was always followed by analysis of a copresent or audio-video interaction.

The analysis was considered complete when, after the fourth pass through the data, no further episodes of breakdowns had been detected. The breakdowns were then tallied and entered into a database. A number of other characteristics of each interaction were collected as well, including the total time taken for each subtask, the number of utterances produced by each participant, the number of utterances appearing in repair sequences for each category of breakdown and so on. While these characteristics did not prove to be useful for the analysis presented in this dissertation, they may be relevant for future analyses; Appendix D presents the complete data record for the analysis.

Table 5.1 presents the data derived by the analysis, showing the number of breakdowns that occurred in each interaction. Each interaction is shown as one row in the table. Interactions are grouped by the environment in which they took place and can be identified by their labels: interactions labeled “FF” took place in the copresent (face-to-face) environment, interactions labeled “AO” took place in the audio-only environment, interactions labeled “AV” took place in the audio-video environment.

Table 5.1: Summary of breakdowns counted in each environment; FF = face-to-face (copresent) environment, AO = audio-only environment; AV = audio-video environment. Each row describes one interaction, detailing the number of breakdowns that occurred in each of the four phases of the task solution process.

	Verbal BD					Cursor BD					Reference BD					Topic BD				
	Subtask#					Subtask#					Subtask#					Subtask#				
Session	**1**	**2**	**3**	**4**	**Ttl**	**1**	**2**	**3**	**4**	**Ttl**	**1**	**2**	**3**	**4**	**Ttl**	**1**	**2**	**3**	**4**	**Ttl**
FF2-FF	9	3	15	21	**48**	0	0	0	2	**2**	2	0	1	0	**3**	0	0	1	2	**3**
FF3-MF	5	4	4	5	**18**	0	0	0	0	**0**	0	1	0	0	**1**	0	0	0	2	**2**
FF4-MM	7	13	9	10	**39**	0	0	0	0	**0**	0	1	2	1	**4**	0	0	2	0	**2**
FF5-MM	11	2	4	5	**22**	0	0	0	0	**0**	0	0	3	1	**4**	1	1	0	1	**3**
FF-Totals					**127**					**2**					**12**					**10**
FF-StDev					**12.3**					**0.9**					**1.2**					**0.5**
AO2-FF	4	1	4	7	**16**	2	1	1	0	**4**	0	1	1	2	**4**	1	1	1	2	**5**
AO3-FF	30	10	10	25	**75**	18	0	8	1	**27**	4	2	1	0	**7**	3	2	2	4	**11**
AO4-FF	11	10	10	32	**63**	2	0	1	1	**4**	2	4	4	1	**11**	2	1	1	1	**5**
AO5-FF	11	4	7	16	**38**	5	1	3	0	**9**	2	1	0	1	**4**	1	0	1	2	**4**
AO-Totals					**192**					**44**					**26**					**25**
AO-StDev					**22.8**					**9.5**					**2.9**					**2.8**
AV2-MM	9	7	15	9	**40**	0	0	3	0	**3**	2	5	0	0	**7**	1	0	2	1	**4**
AV3-MF	9	23	13	22	**67**	11	5	8	0	**24**	1	4	4	3	**12**	3	5	3	1	**12**
AV4-FF	5	1	5	2	**13**	0	0	0	0	**0**	1	1	1	0	**3**	0	0	1	0	**1**
AV5-FF	1	0	8	4	**13**	2	0	3	0	**5**	5	0	1	0	**6**	1	0	4	2	**7**
AV-Totals					**133**					**32**					**28**					**24**
AV-StDev					**22.4**					**9.4**					**3.2**					**4.1**

To provide a sense for how breakdowns were distributed within interactions, Table 5.1 details the number of breakdowns that occurred in each of the four major subtasks (see Chapter III) that participants had to address in order to accomplish the overall CVCK task. Briefly, subtasks one and three were primarily construction-oriented, involving assembly or modification of the cardiovascular loop whose behavior participants were working to understand; subtasks two and four were primarily analytic, requiring participants to run the simulator to answer a series of questions about the physiological behavior of the cardiovascular construct.

An examination of Table 5.1 reveals that there were substantial differences in the number of breakdowns documented in the three communication environments. To one extent or another, copresent interactions showed a lower total incidence of breakdown in all four categories than either of the two technologically-mediated environments. For both Topic and Reference breakdown, the total number of breakdowns documented in audio-only and audio-video interactions was roughly twice the number documented in copresent interactions. The difference was even more striking for Cursor turntaking breakdowns — participants in audio-video and audio-only interactions exhibited approximately *twenty times* more Cursor turntaking breakdown than copresent participants. Only in the category of Verbal turntaking breakdown were the differences between environments somewhat less compelling, with copresent and audio-video interactions showing roughly similar numbers of breakdowns, while audio-only interactions showed a slightly higher total. At the same time, the differences between the audio-only and audio-video environments were generally small — with the exception of Verbal turntaking breakdown, the total number of breakdowns observed in these two environments was roughly the same.

The fact that the number of breakdowns documented in the copresent environment was lower in all four categories clearly suggests that copresent participants were more adept at organizing their interactions than were participants in the audio-only or audio-video environments. Another useful measure of expertise (Card, Moran, & Newell, 1983) is the *variability* in the level of observed performance — for any given task, a group of experts will tend to exhibit a stable and similar level of performance, while the performance of non-experts will vary widely between individuals. Looking at the variability (as measured by standard deviation) in the number of breakdowns per interaction for each environment, it is evident that the amount breakdown per interaction was much more consistent in the copresent environment than in the two technologically-mediated environments. For example, the standard deviation in the number of Cursor turntaking breakdowns per interaction was roughly nine times higher in the two technologically-mediated environments than in the copresent environment; though not quite as dramatic, the variability in the remaining three categories of breakdown was substantially lower in the copresent environment as well. In conjunction with the lower overall incidence of breakdown in copresent interactions, these observations support the conclusion that participants were relatively "expert" at copresent interaction, but decidely non-expert at technologically-mediated interaction.

Finally, it is important to point out that there are no obvious trends in the distribution of breakdowns within interactions in any of the three environments. In particular, the number of breakdowns that occurred near the end of interactions (i.e. during the latter subtasks) was not consistently lower than the number that occurred during the earlier subtasks for interactions in any of the three environments. The absence of such decreasing tendencies in number of breakdowns indicates that the communicative performance of participants in technologically-mediated interactions did not noticeably improve as they gained experience interacting in the environment. Though this does not rule out the possibility that participants in technologically-mediated interactions might *eventually* develop compensatory organizational mechanisms to reduce the number of breakdowns they experience, it does suggest that such adaptations are non-trivial and can not be accomplished in the short term.

In sum, a preliminary review of the quantitative results shows that, with the exception of Verbal turntaking breakdown, the total number of breakdowns that occurred in the copresent environment was substantially lower than in the audio-only or audio-video environments. By contrast, the differences between the audio-only and audio-video environments were relatively small, with both environments showing similar total amounts of breakdown in all categories except Verbal turntaking. These results are graphically summarized in Figure 5.1.

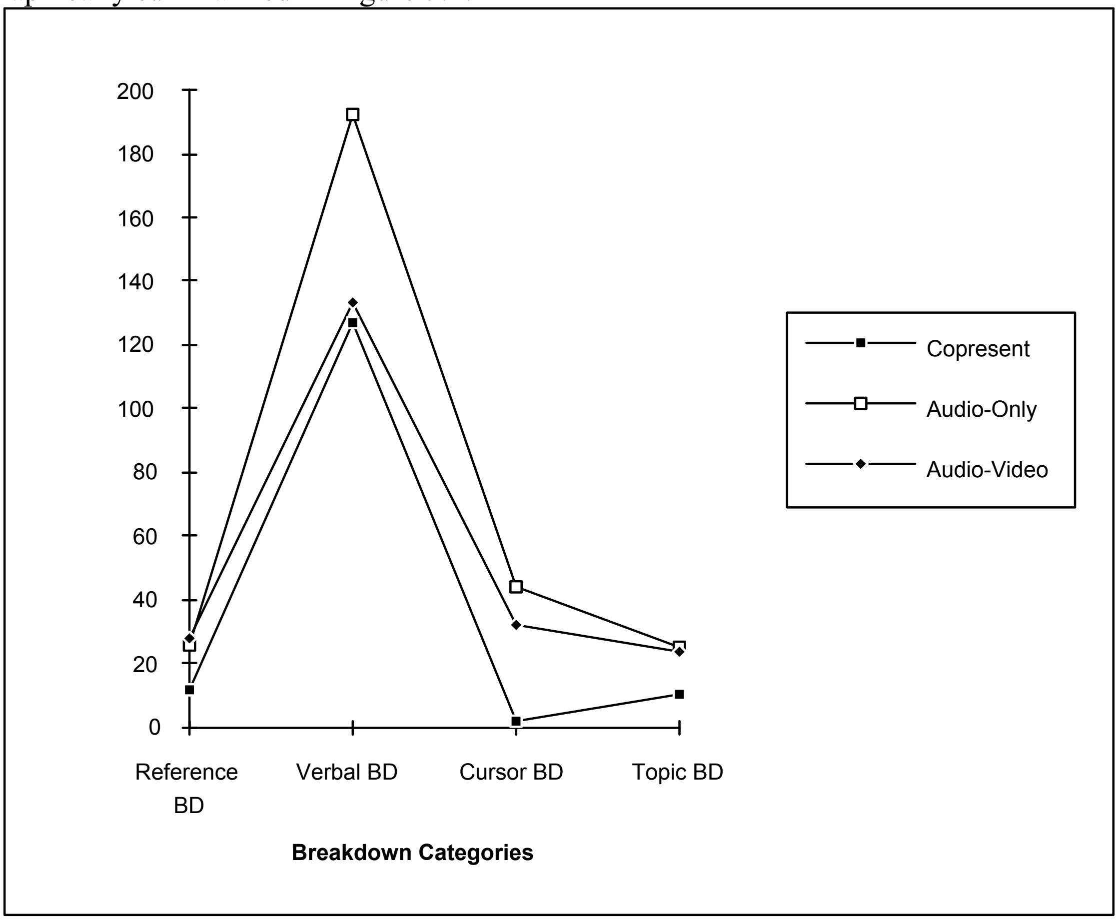


Figure 5.1: Comparative overview of total number of breakdowns observed in each category for all interactions in each of the three communication environments.

Though these differences in the *total* number of breakdowns that occurred in each environment are quite compelling, a closer examination of the data presented in Table 5.1 shows that there was considerable variation in the amount of communicative breakdown that occurred between *individual pairs* of participants, both within and between communication environments. For instance, several audio-only and audio-video interactions showed less Verbal turntaking breakdown than some copresent interactions; in some audio-video interactions there was less Topic, Cursor turntaking or Reference breakdown than in certain copresent interactions.

The question raised by this local variability is whether the differences in total amount of breakdown apparent in Figure 5.1 can be attributed to variations in the amount of breakdown that are bound to exist in any sampling of human performance, or whether they reflect consistent, underlying differences in the level of communicative support provided by each of the three communication environments. In the following section, this question is formally answered by applying nonparametric statistical techniques to test whether variation in the amount of breakdown observed for each environments is statistically significant.

5.2 Statistical Analysis

The total number of breakdowns that occurred in each of the three environments was compared for each of the four categories of breakdown using nonparametric statistical techniques. Though the power of nonparametric techniques is somewhat lower than that of parametric techniques commonly used in the social sciences, two features of the breakdown data analyzed ruled out the use of such parametric techniques for this analysis:

1. Nature of Data. Most parametric techniques for testing for significant differences require that sample data be of at least interval-ratio strength, since these techniques are based on the numerical analysis of the magnitude of differences between samples. Though the data (i.e. communicative breakdown) in this analysis are represented by integer values, it is unrealistic to assume that these values fall on an

interval-ratio scale, given the qualitative nature of the analysis used to generate these data. For instance, it is not clear that the amount of communicative trouble evidenced by 20 breakdowns in one interaction is *exactly twice* the amount of communicative trouble evidenced by 10 breakdowns in a different interaction. More generally, the inherent subjective component of the analytic process used to identify episodes of communicative breakdown implies that statistical analysis should not rest on fine-grained distinctions in the number of breakdowns that occurred in interactions; it is simply unrealistic to assume that data derived from a qualitative analysis are as objective and consistent as those yielded by the mechanical instrumentation typically relied on in the physical sciences (Siegel & Castellan, 1988) . Consequently, the breakdown data collected in this study was considered to be *ordinal* in nature.

2. Sample size. Another limitation of the data generated by the Breakdown Analysis presented in this study is the relatively small sample size (i.e. N=4). Because of the substantial amount of time and effort required to expose breakdowns that occur in communicative interaction, it is impractical to examine large numbers of interactions in each environment. While most parametric techniques require large sample sizes to produce meaningful results, many nonparametric techniques have been developed specifically for situations in which samples size are very small.

In sum, nonparametric techniques were selected for the statistical analysis of the breakdown data collected in this study because of the relatively small sample size and the ordinal nature of the data. Rather than focusing on the magnitude of differences between samples, these techniques focus on exposing statistically significant patterns in rank ordering of samples. This amounts to a fairly conservative approach to evaluating differences between environments, emphasizing the consistency of differences while ignoring their scale — only very consistent patterns of difference will be deemed to be significant. It is this conservative character that allows nonparametric techniques to provide a high level of certainty, even for small sample sizes.

5.2.1 Statistical Analysis: Details

The following paragraphs formally describe the statistical analysis performed on the breakdown data.

5.2.1.1 Comparisons performed

The three communicative conditions under which interactions took place — copresent, audio-only, and audio-video — represent the independent variables in the statistical analysis; the four categories of breakdown represent the dependent variables. Interactions were grouped by the environment in which they took place and these groups compared based on the number of breakdowns observed in each category. A total of 12 statistical tests were performed, comparing the three environments on each of the four categories of breakdown.

5.2.1.2 Hypotheses tested

The null hypothesis, H_0, and the alternative hypothesis, H_1, were different for each of the 12 comparisons performed. Table 5.2 summarizes the H_0 and H_1 for each comparison, organizing them by which environments were being compared.

Table 5.2: Summary of experimental hypotheses for the 12 statistical comparisons performed.

Environments	Hypotheses
Copresent compared to Audio-Only	**H_0**: the amount of {x} breakdown was the same in Copresent interactions as in Audio-only interactions. **H_1:** the amount of {x} breakdown was larger in Audio-only interactions than in Copresent interactions. where x={Verbal turntaking, Cursor turntaking, Reference, Topic}
Audio-only compared to Audio-video	**H_0:** the amount of {x} breakdown was the same in Audio-only interactions as in Audio-video interactions. **H_1:** the amount of {x} breakdown was larger in Audio-only interactions than in Audio-video interactions. where x={Verbal turntaking, Cursor turntaking, Reference, Topic}
Audio-video compared to Copresent	**H_0:** the amount of {x} breakdown was the same in Copresent interactions as in Audio-video interactions. **H_1:** the amount of {x} breakdown was larger in Audio-video interactions than in Copresent interactions. where x={Verbal turntaking, Cursor turntaking, Reference, Topic}

In general, the null hypotheses assert that two environments do not differ significantly in the amount of breakdown observed in each of the four categories, while the alternative hypotheses assert that the number of breakdowns is significantly higher in one of the environments than in the other.

5.2.1.3 Statistical Test

The Mann-Whitney *U* test was used to perform the statistical comparisons. This test is the most powerful nonparametric test available, used as an alternative to the parametric *t* test when the sample data is ordinal in scale. Importantly, the Mann-Whitney *U* test is particularly well-suited for situations involving small sample sizes.

5.2.1.4 Assumptions

The Mann-Whitney test assumes that samples are independent. This criterion was met by the data collected in this study, since no subject was allowed to participate in more than one interaction.

5.2.1.5 Significance Level

The level of significance was set at $\alpha = 0.1$. Though this level of significance is slightly higher than that traditionally adopted in parametric statistical studies, it is not at all unusual for nonparametric analyses (Siegel & Castellan, 1988) . More generally, choosing a level of significance for behavioral studies is essentially a matter of judgment, requiring the analyst take into account the nature of the data and the domain, as well as the ultimate goals of the analysis. Because the statistical analysis of breakdown does *not* constitute the final result of the Breakdown Analysis presented in this dissertation, but rather serves to focus a subsequent qualitative examination of participants' communicative behavior, a slightly less stringent level of significance was justified. In particular, the level of significance of $\alpha = 0.1$ was chosen to accentuate consistent patterns in a relatively small sample of data, while still providing a high degree of statistical certainty. The actual probabilities associated with each test are noted alongside the results presented in upcoming sections, making apparent the relationship between the level of significance chosen and the conclusions drawn from the analysis.

5.2.1.6 Sampling Distribution

The size of the samples compared in each of the 12 statistical comparison was identical: $n_1=n_2=4$, corresponding to the four interactions that took place in each communication environment. The variation of the Mann-Whitney test used for this analysis was designed specifically for such small sample sizes; the exact probabilities associated with the occurrence under H_0 of various *U* values reported here were drawn from tables given in (Siegel & Castellan, 1988) .

5.2.1.7 Rejection Region

Since the alternative hypothesis (H_1) for each comparison states a particular *direction* of difference, the region of rejection for all 12 comparisons is one-tailed and consists of all *U* values which are so small that their probability of occurrence under H_0 is equal to or less than $\alpha = 0.10$.

5.2.2 Results

The breakdown data collected for the various interactions analyzed were compared using the statistical method described in the previous section. The results are summarized in Table 5.3.

As indicated in Table 5.3, the comparison of breakdown frequency reveals that some of the differences in frequency of breakdown implied in Figure 5.1 are significant while others are not. Specifically, the results of the statistical analysis were as follows:

1. Verbal turntaking breakdown. No significant differences in the number of Verbal turntaking breakdowns were found between any of the three environments.

2. Cursor turntaking and Reference breakdown. The number of Cursor turntaking and Reference breakdowns was significantly lower for interactions in the

copresent condition than for those in either the audio-only condition or the audio-video condition. No significant differences in the amount of either Cursor turntaking or Reference breakdown were found between the audio-only and audio-video environment.

3. Topic breakdown. The number of Topic breakdowns was significantly lower for copresent interactions than those occurring in the audio-only condition. No significant differences in the amount of Topic breakdown were found between copresent and audio-video interactions, or between audio-video and audio-only interactions.

Table 5.3: Results of comparison of frequency of breakdown between environments in each of the four categories, using Mann-Whitney *U* test (α=0.1, N=4). The p_{Ho} values give the probability of Type I error (falsely rejecting H_0 when it is true) associated with each comparison.

	VERBAL	CURSOR	REFERENCE	TOPIC
Copresent compared to Audio-only	No significant difference (U=6; p_{Ho}=0.343)	Copresent has significantly FEWER breakdowns (U=0; p_{Ho}=0.014)	Copresent has significantly FEWER breakdowns (U=2; p_{Ho}=0.057)	Copresent has significantly FEWER breakdowns (U=0; p_{Ho}=0.014)
Copresent compared to Audio-video	No significant difference (U=7; p_{Ho}=0.443)	Copresent has significantly FEWER breakdowns (U=2.5; p_{Ho}=0.064)	Copresent has significantly FEWER breakdowns (U=2.5; p_{Ho}=0.064)	No significant difference (U=4; p_{Ho}=0.171)
Audio-only compared to Audio-video	No significant difference (U=5; p_{Ho}=0.243)	No significant difference (U=5; p_{Ho}=0.243)	No significant difference (U=7.5; p_{Ho}=0.5)	No significant difference (U=7.5; p_{Ho}=0.5)

In sum, copresent interactions were generally less prone to breakdown than technologically-mediated interactions. Aside from Verbal turntaking breakdown, where there was no difference between any of the three environments, the only comparison which did *not* reveal a significantly lower amount of breakdown in the copresent condition was the comparison of Topic breakdown between copresent and audio-video interactions. Note however that, though the differences revealed in this comparison

(U=4, p_{Ho}=0.171) do not quite meet the level of significance established earlier, they are clearly more substantial than in any of the other comparisons for which no significant difference was found.

Based on the statistically differences in the amount of breakdown revealed by the analysis presented above, the following conclusions regarding the relative communicative efficacy of the three environments compared may be drawn:

Copresent versus Audio-only interaction.

> The communicative efficacy of copresent interactions was higher than that of interactions in the audio-only condition. The amount of breakdown in copresent interactions was significantly lower than in audio-only interactions in three out of four categories; in no category was the amount of breakdown significantly lower for audio-only interactions than for copresent ones.

Copresent versus Audio-video interaction.

> The communicative efficacy of copresent interactions was higher than that of interactions in the audio-video condition. The amount of breakdown in copresent interactions was significantly lower than in audio-video interactions in two out of four categories; in no category was the amount of breakdown significantly lower for audio-video interactions than for copresent ones.

Audio-only versus Audio-video interaction.

> There was no difference in the communicative efficacy of interactions occurring in the audio-only and the audio-video conditions. No significant difference in the amount of communicative breakdown was found between audio-only and audio-video interactions in *any* of the four categories of breakdown analyzed.

In sum, this analysis shows that the communicative efficacy of copresent interactions was substantially higher than that of technologically-mediated interactions, clearly demonstrating that the two technologically-mediated communication environments examined in this study are not functionally equivalent to the copresent condition — participants working in the two technologically-mediated environments had significantly more difficulty establishing and maintaining shared interpretations of their collaborative interaction than copresent participants.

5.2.2.1 Discussion: Invalidating the Bandwidth Assumption

One reason that audio-only and audio-video environments were chosen for comparison in this analysis was to empirically test the Bandwidth Assumption, which has been tacitly used to rationalize the design of many current systems. As discussed in Chapter I, the Bandwidth Assumption asserts that the communicative efficacy of interactions that take place in a given technologically-mediated environment is directly related to the bandwidth of the connection that the environment provides between interacting participants — higher bandwidths necessarily lead to more effective interactions than lower ones. Clearly, the addition of a video channel makes the bandwidth of the connection provided in the audio-video condition substantially higher than the bandwidth provided in the audio-video condition. Accordingly, the Bandwidth Assumption predicts that interactions in the audio-video environment should exhibit significantly higher communicative efficacy. The results of this analysis, which showed that the communicative efficacy of audio-video interactions was essentially the same as that of audio-only interactions constitutes strong empirical evidence that the Bandwidth Assumption is not a reliable basis for characterizing the extent to which technologically-mediated environments support the communicative endeavors of users.

It is important to emphasize that the results of this analysis do not mean that supporting higher bandwidth can*never* improve communicative efficacy and that, therefore, designing high-bandwidth environments is pointless. Rather, the results indicate that there is not *necessarily* a direct relationship between bandwidth and communicative efficacy. As we shall see in the upcoming analysis presented in Chapter VI, it is not the total volume of information that matters, but the communicative resources embodied in and made available through augmentations in bandwidth; if improvements in bandwidth can provide access to these resources, they might be very valuable indeed.

5.2.2.2 Discussion: Comparison to Other Empirical Studies.

It is interesting to compare the results yielded by this analysis to the results yielded by other empirically-oriented investigations of technologically-mediated interactions. As discussed in Chapter I, existing approaches can be classified by the metric they use to characterize the communicative efficacy of interactions: user satisfaction, quality of work, and task-activity structure. Table 5.4 summarizes the results yielded by these approaches.

Existing studies of user satisfaction (Isaacs, Morris et al., 1995; Olson, Olson et al., 1995; Tang & Isaacs, 1992; Tang, Isaacs et al., 1994) have found that users overwhelmingly prefer copresent to technologically-mediated interaction when given a choice. However, these same studies also show that users consistently perceive interaction in audio-video environments to be more natural and satisfying than interaction in an audio-only environment. In this way, these studies suggest that copresent interactions have the highest communicative efficacy, followed by audio-video interactions and audio-only interaction.

Table 5.4: Comparison of results of other empirical studies of copresent, audio-only and audio-video environments.

	Copresent versus Audio-only	Copresent versus Audio-video	Audio-only versus Audio-video
User Satisfaction	Copresent has higher preference	Copresent has higher preference	Audio-video has higher preference
Quality of work	Copresent has higher quality	No Difference	No Difference
Task-structure	Audio-only has more organizing and planning	Audio-video has more organizing and planning	No Difference

By contrast, comparison of the quality of work produced by participants (Olson, Olson et al., 1995) revealed a significant difference only between audio-only and copresent interactions, implying that the difference in communicative efficacy between the three environments as a whole is very small. Finally, studies of the task-activity structure of interactions occurring in different environments have revealed significant differences in the way that participants go about accomplishing their collaborative goals in copresent and technologically-mediated interactions. For example, Olson et al. (1995) have shown that participants in technologically-mediated environments spend significantly more time organizing and planning their interaction, as well as clarifying what they meant, than do copresent participants; no significant difference was found in the structure of interactions in audio-only and audio-video conditions. Several less formal analyses (Tang, 1991; Tang, Isaacs et al., 1994; Tatar, 1989) of task activities have documented other differences in copresent and technologically mediated interaction as well. Though such task-activity studies are generally aimed more at articulating differences than at comparing communicative efficacy, the nature of these differences implies that copresent interaction is more effective than technologically-mediated interaction.

5.3 Summary: Differences in Communicative Efficacy

The central research issue explored in this dissertation is to what extent interaction in technologically-mediated communication environments can be considered "functionally equivalent" to copresent interaction; is the communicative efficacy of interactions occurring in technologically-mediated environments just as high as for similar interactions between copresent participants? The goal of the second, quantitative phase of the Breakdown Analysis, the results of which were presented in this chapter, was to begin to answer this question by statistically comparing the amount of breakdown that occurred in interactions that took place in the three communication environments evaluated in this study.

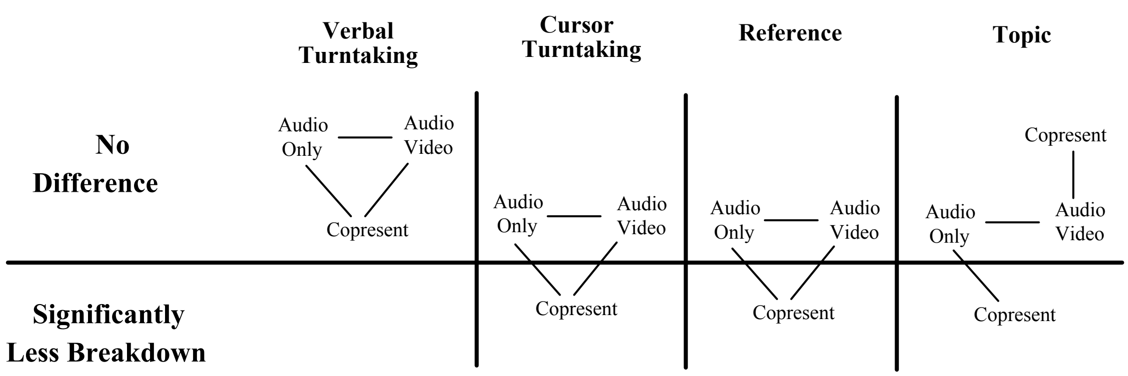


Figure 5.2: Summary of statistical comparison of total breakdown observed in the three communication environments.

After using the criteria developed in Chapter IV to determine the total number of Verbal turntaking, Cursor turntaking, Reference, and Topic breakdowns that occurred in each interaction, nonparametric techniques were used to test for significant differences in the amount of breakdown observed between interactions that took place in the copresent, audio-only, and audio-video conditions. The results of this analysis are summarized in Figure 5.2.

As indicated in Figure 5.2, interactions in the copresent condition were found to have significantly fewer breakdowns than technologically-mediated interactions in almost all cases; the only exceptions were Verbal turntaking breakdown, where no significant differences were found between any of the three environments[1], and Topic breakdown, where no significant differences were found between copresent interactions and those in the audio-video condition.

Based on these results, the following conclusions were drawn with respect to the relative communicative efficacy of interaction in the three communication environments evaluated:

1. The communicative efficacy of copresent interactions was substantially higher than that of interactions in the two technologically mediated environments.

2. There was no difference in communicative efficacy between the audio-only and audio-video conditions.

These results clearly imply that the two technologically-mediated environments fall short of their implicit design goal of providing simulacrums of copresent interaction; participants were not able to accomplish their communicative goals in these environments as effectively and efficiently as when they were physically copresent. Moreover, the availability of a video channel did not appear to significantly enhance the efficacy of interactions in the audio-video condition.

Though these results provide compelling evidence that technologically-mediated interaction is not a functional substitute for physical copresence, they shed no light on *why* this is the case. In order to inform the design of future technologically-mediated environments, we need to somehow understand how differences in communicative efficacy revealed by the quantitative analysis of breakdown are related to the design of the two technologically-mediated environments examined in this study. That is, how can observed differences in communicative efficacy be rationalized? In the next chapter, we present the results of the third and final phase of the Breakdown Analysis, which works to answer this question by qualitatively examining individual episodes of breakdown that occurred in the audio-only and audio-video environments.

5.4 Notes

[1] Though the focus of this analysis is primarily on exposing and explaining *differences* in the number of breakdowns that occurred in the three environments, it is, in principle, equally interesting to ask why there were no significant differences in Verbal turntaking breakdown between environments. We return to this issue in Chapter VII, in the context of discussing future work.

CHAPTER VI

RATIONALIZING BREAKDOWN

The quantitative comparison of communicative breakdown presented in Chapter V clearly implies that the three communication environments examined in this study differed significantly in the extent to which they supported the communicative endeavors of participants. Specifically, the analysis revealed that copresent interactions showed a significantly lower incidence of breakdown than audio-only interactions in three out of four categories of breakdown, and a significantly lower incidence of breakdown than audio-video interactions in two out of four categories; in no category was the incidence of breakdown significantly lower for technologically-mediated than for copresent interaction. Collectively, these results suggest that the communicative efficacy of copresent interactions was consistently higher than that of interactions in the audio-video or audio-only environments. Finally, the fact that there were no significant differences in the number of breakdowns in any of the four categories between audio-only and audio-video interaction implies that there was no difference in communicative efficacy between the audio-only and audio-video conditions.

While these results are interesting, they are of limited utility from a practical standpoint. For designers of technologically-mediated environments, knowing that the communicative efficacy supported by a given technologically-mediated environment is not equivalent to copresent interaction is less important than understanding why this deficiency exists. That is, how can the failure of the technologically-mediated environments to support interactions that are functionally equivalent to copresent interaction be attributed to the physical characteristics of the environment? More specifically, the results yielded by the quantitative analysis of breakdown frame two questions relevant to the design of future technologically-mediated environments:

1. How do the observed deficiencies (compared to copresent interaction) in the communicative efficacy of interaction in the audio-only and audio-video environments arise from the design of these environments?

2. Why was there no difference in communicative efficacy between the audio-video and audio-only conditions, despite the fact that the audio-video condition was obviously a more sophisticated environment? In particular, why did the availability of a video channel not significantly reduce the amount of breakdown experienced in this environment?

The goal of this chapter is to answer the above questions by presenting the results of the third and final phase of the Breakdown Analysis, in which breakdowns that occurred in technologically-mediated interactions were subjected to an intensive qualitative analysis, in an effort to rationalize the differences in communicative efficacy revealed in Chapter V. Clearly, addressing this issue requires somehow constructing causal relationships between certain physical characteristics of the two technologically-mediated environments and the communicative breakdowns that occurred during the interactions that took place in those environments. At the same time, the fact that communicative breakdown is an intrinsically epistemic phenomenon means that the nature of the causal explanations we can expect to construct and the analytic approach taken to generate these explanations will be constrained by the epistemological foundations of the analysis.

The following section lays out the analytic framework for the chapter, first establishing an epistemological foundation for the investigation and then describing the analytic procedures. Subsequent sections present, respectively, the results of the analysis for Cursor turntaking, Reference, and Topic breakdowns, working to explain the significant differences in the incidence of these breakdowns between copresent and technologically-mediated interactions revealed in Chapter V. Finally, the closing sections of the chapter discuss the results and present further evidence to arrive at overall conclusions regarding the communicative efficacy of technologically-mediated interaction.

6.1 Analytic Framework

In traditional scientific domains that are concerned with physical phenomena, "explaining" some observed phenomenon means positing a context-independent causal relationship between certain abstract characteristics of the situation and the phenomenon. For instance, the observation that a ball drops to the ground when released can be explained by the fact the ball and the earth both have mass, and that a gravitational attraction exists between all masses. In this way, scientific explanation amounts to articulating the ways in which a particular observed behavior can be seen as the instantiation of certain context-independent "laws" of physical behavior defined by a comprehensive underlying model of the physical world.

The fact that communication is an epistemic process, defined by cognitive rather than physical phenomena, makes the enterprise of rationalizing communicative breakdown fundamentally different from the explanation of physical phenomena. As discussed in Chapter II, the epistemology of Situated Action asserts that shared understanding is constructed uniquely in each situation through participants' contextual interpretation of each other's communicative displays. In particular, Situated Action denies the existence of context-independent interpretive rules, or "scripts," as a basis for organizing interaction and constructing its significance. This commitment to communication as a dynamic, locally-negotiated process rules out context-independent explanations for communicative breakdown; the fact that there exists no universal model of communicative behavior makes it impossible to explain domain phenomena (i.e. communicative breakdown) in context-independent terms.

Even as it rules out context-independent explanations of breakdown, however, the conception of how shared meaning arises under Situated Action suggests a more appropriate basis for rationalizing differences in the amount of communicative breakdown observed in different environments. Specifically, the fact that communication is characterized as the dynamic, situated interpretation of the communicative displays, e.g. utterance, gaze, gesture, of conversational partners makes it is reasonable to presume that communicative breakdown will be *more likely* to occur in contexts in which access to these evidentiary resources is somehow constrained. This idea is illustrated in Figure 6.1.

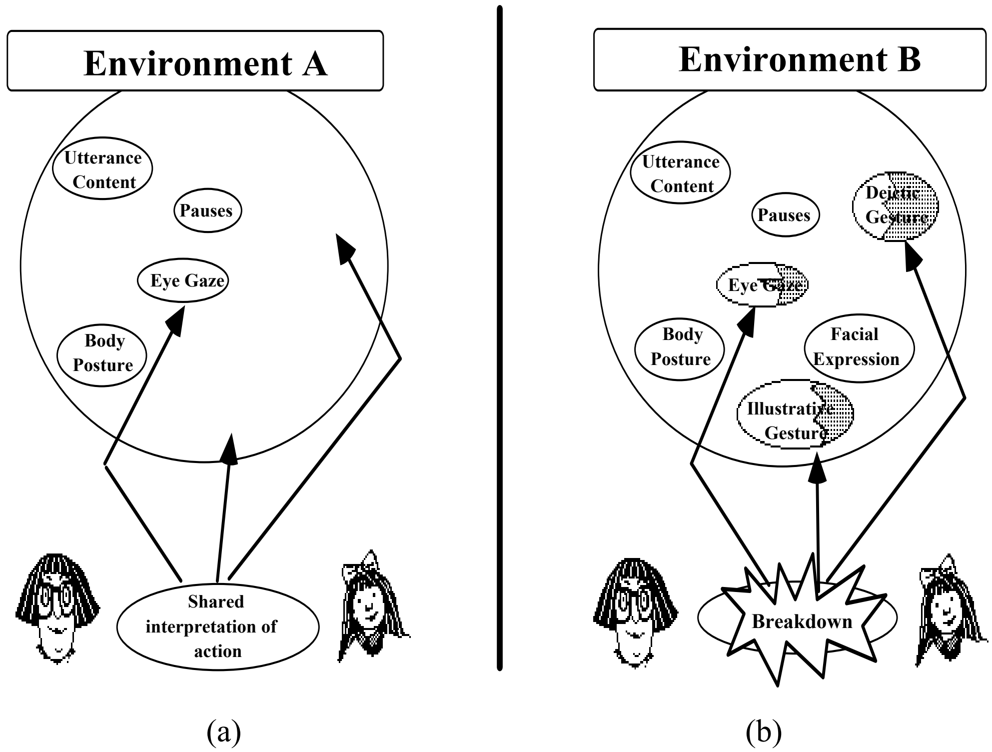


Figure 6.1: Communicative breakdown is more likely in environments (b) in which access to communicative resources is somehow constrained than in those (a) in which it is not.

As suggested in Figure 6.1, participants will be more likely to succeed[1] at constructing shared interpretations of action when they have unrestricted access to the full range of each other's verbal and nonverbal communicative displays. Because inferring a partner's interpretation of action is based on the contextual interpretation of his or her communicative displays, the robustness of this process is directly related to the amount and quality of the evidence available. When access to these communicative resources is somehow restricted by the environment in which participants are interacting (Figure 6.1b), this evidentiary process is effectively crippled, resulting in a greater likelihood that participants will fail to maintain shared interpretations of action, resulting in a communicative breakdown.

The presumption that the likelihood of communicative breakdown is directly related to the communicative resources accessible to participants establishes a *probabilistic* causal framework for the analysis of the differences in communicative efficacy revealed in Chapter V: the communicative efficacy of copresent interactions was higher than that of technologically-mediated interactions because the two technologically-mediated environments somehow constrained participants' access to certain communicative resources, thereby increasing the likelihood of communicative breakdown as participants struggled to maintain intersubjectivity in an impoverished evidentiary context.

6.1.1 Overview of Analytic Process

As discussed in the previous section, the only way to rationalize the observation that the communicative efficacy in technologically-mediated interactions was significantly lower than that of copresent interaction is by identifying constraints on certain communicative resources imposed by the technologically-mediated interactions. Accordingly, analytic attention was focused on first characterizing the communicative resources that participants used to organize their interactions in general, and then working to expose regularities in the communicative resources that were available to participants in technologically-mediated interactions at the time that breakdowns occurred; if it can be established that breakdowns consistently occur when participants relied on certain types of communicative resources, then this strongly implies that these resources were somehow rendered inaccessible by the electronic environment. This two step analytic process is graphically summarized in Figure 6.2:

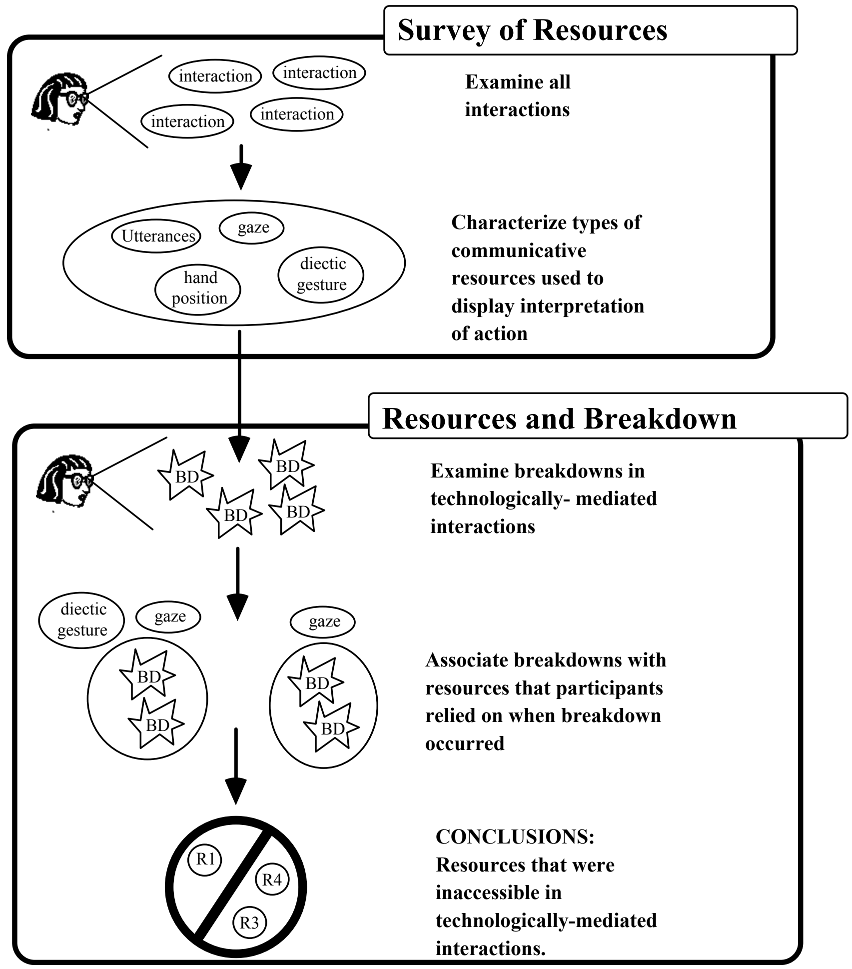


Figure 6.2: Overview of the two stage analysis of communicative resources available to participants during episodes of breakdown.

In the first step of the analysis, interactions taking place in all three environments were examined to characterize the various communicative resources used by participants to organize their cursor, reference and topic management activities. By articulating the range of communicative displays used to organize these three aspects of interaction, this preliminary analysis established the basis for articulating consistent relationships between communicative resources and breakdown in the second step of the analysis.

In the second step of the analysis, breakdowns in each of the three categories in which significant differences in communicative efficacy were found — i.e., Cursor turntaking, Reference, and Topic — were collected and analyzed to expose underlying regularities in the communicative resources that were available to participants when breakdowns occurred. By revealing that breakdowns in technologically-mediated interactions were consistently related to certain kinds of communicative displays (and the absence of others), these observations provide a basis for concluding that access to these displays was somehow constrained in technologically-mediated environments. Because of this constrained access, the overall likihood of breakdown was increased, resulting in the lower communicative efficacy observed for these environments. To further support these conclusions, copresent interactions were examined as well to establish that copresent participants were able to effectively utilize the displays in question to inform their interactions.

The following three sections present, respectively, the results of applying this qualitative analysis to rationalize the differences in Cursor turntaking, Reference and Topic breakdown exposed by the earlier quantitative comparison. Each section begins by characterizing the various communicative displays that participants made available as resources for organizing the aspect of their interaction in question, followed by a discussion of consistent relationships between communicative resources and the breakdowns that occurred in technologically-mediated interactions. Section 6.5 sums up the results of the analysis, bringing together the patterns of observations made for each category of breakdown to highlight common themes of resource constraint, and relating these constraints to the design of the audio-only and audio-video environments.

6.2 Rationalizing Cursor Turntaking Breakdown

Because there was only a single cursor available in the shared workspace, participants had to maintain a continuous sense of whose turn it was to use the cursor throughout the interaction. Cursor turntaking breakdowns were defined by the failure to adequately manage this organizational process, resulting in confusion over whose turn it was to use the shared cursor; such confusion was predominantly evidenced by simultaneous attempts to control the cursor, resulting in the erratic behavior of both the cursor and the CVCK simulator.

6.2.1 Resources for Cursor Management

The examination of cursor management in interactions in all three environments revealed that participants relied on a variety of verbal and nonverbal displays to negotiate access to the shared cursor. Nonverbal displays included position of the hand with respect to the mouse, movement of the hand towards the mouse, and manipulation of the mouse and shared cursor. Verbal displays were also used extensively, either to request control over the shared cursor, or to explicitly signal the end of a turn at control and pass control to a partner. As a framework for analysis, two distinct approaches to cursor management were identified, based on the type of communicative displays (i.e. resources) that participants relied on to regulate access to the shared cursor: *verbally-regulated* cursor management and *nonverbally-regulated* cursor management. As implied by the name, episodes of verbally-regulated cursor management were characterized by the use of some sort of verbal negotiation over cursor control, during which the issue of who was controlling the cursor was explicitly addressed. By contrast, nonverbally-regulated cursor management was completely tacit, with participants relying entirely on nonverbal displays to regulate access to the shared cursor. The following sections discuss each of these cursor management strategies in more detail.

6.2.1.1 Verbally-Regulated Cursor Management

Exchanges in which participants relied on verbally-regulated cursor management to negotiate control over the cursor were characterized by explicit verbal discussion of cursor control. For instance, consider the following segments:

•	R: ohh-h-hit doesn't move ((2)) R: oh there= M: =okay R: ok*ay:::* M: (you *kego*) ((3))	2- M aligns L2 and drops it (1.0) 3- Both gaze at LB, then R then M raise almost in synch and watch R rolls cursor to L in pallete (1.5)	

Segment: AO4p5

•	R: so::: (.6) this[3] is towards? or::: [M: so:: (.8) R: this (arrow here) [M: .hhh uuumm thats what I was figuring so:: wh::ich one is the (heart),-ooohwell I= [R: okay M: =thi::n::k- you know I think-- [4]can I --- move that cursor?= R: =sure ((5))	3- R motions right to left on top of top G2g axis arrow (animating it essentially) 4- M grabs mouse 5- R jerks hand from mouse and hits mic for loud crunch just as M starts speaking (.5)	

Segment: AO4p34

•	((2)) R: awright- [3]its my turn to play ((4)) M: wait! wait! wait! --- [5]lemme click it off [R: (I doubt it!) (.9) M: [6]see its not in the middle ((7)) M: kay: ------ your turn [R: ohh yea::h there you go --- [8]my turn	2- M drags out new C and positions as C1 (2.3) 3- R moves hand to mouse, M still positioning C 4- M drops C to place as C1 5- M clicks C1, R makes shrugging gesture 6- M grabs C1 and starts to align it 7- M carefully aligns C1, then drops it (1.8) 8- R jerks cursor over to pallete, M drops to LB	

Segment: AV3p4

The exchange shown in segment AO4p5 presents a very straightforward example of verbally-regulated cursor management, in which one participant produces a verbal display (i.e. "you kego") to explicitly mark the end of her turn at cursor control, and pass control of the cursor to her partner. The exchange shown in segment AO4p34 is similar, except that in this case it is the passive participant who verbally requests control over the cursor from her partner, who is currently in control of the cursor.

In the exchange presented in segment AV3p4, negotiation over cursor control is somewhat more extensive. The exchange begins with R verbally expressing his intention to take control of the shared cursor, creating an explicit opportunity for M to extend her turn at control until she has finished with the task at hand before verbally passing control of the cursor to R. This exchange illustrates an important feature of verbally-regulated cursor management, namely, that it makes explicit participants' interpretations of whose turn it is to control the shared cursor next and when that turn at control is to begin or end, essentially introducing the issue of cursor control as a distinct subtopic of discussion. By bringing the issue of cursor control into the conversational foreground in this way, verbally-regulated cursor management provides participants with the opportunity to expose and resolve nascent confusions over cursor control before they result in Cursor turntaking management breakdown.

6.2.1.2 Nonverbally-Regulated Cursor Management

An obvious disadvantage of verbally-regulated cursor management is that it is relatively cumbersome, with each turn at control explicitly negotiated through verbal discussion. In this sense, it is very much like having to append the word “over” to mark the end of verbal turns, as was done for many years in the world of radio communications. When using the nonverbally-regulated cursor management strategy, participants avoided this organizational “overhead” by relying solely on tacit, nonverbal displays to regulate access to the shared cursor. That is, participants simply took control of the cursor in an opportunistic fashion, relying on nonverbal resources rather than explicit verbal negotiation to infer that the cursor was currently available. As illustrated in the following segments, several kinds of nonverbal behavior may serve as tacit evidence that an opportunity to take control of the shared cursor is at hand:

•	((1)) R: so *we just like - click it over there?* ((2)) M: *shu:::*[3]*r* R: wu::: lets see[4] what happens ((5)) R: <u>ho</u>::ly smokes	1- R gazes WS, M gazes LB, then R drops to LB, speaking (1.8) 2- Both stare LB (2.0) 3- R snaps to WS and grabs mouse 4- R rolls cursor noisily over to pallete 5- R puts cursor on C and drags a new C into the WS and holds it there. M is still gazing LB, tracing sentences with pen (3.7)	5- new C appears in WS

Segment: FF4p1

•	R: how do we get rid of[2] this? ((3)) M: try the[4]::: ((5)) R: tr-<u>hi</u> sump'mm [M: sump'mm (.7) R: hummm[6]mmm ((7))	2- R points and clicks on G2a graph with cursor 3- R rolls cursor back and forth, M glances towards the table (1.7) 4- M grabs his mouse 5- R rolls mouse some more, then pulls back hand and shrugs as he speaks (1.2) 6- M clicks on G2a 7- M hesitates, then drags G2a over to the biowaste (4.2)	2- G2a highlights 7- biowaste highlights as it is contacted.

Segment: FF4p14

•	(.8) M: **uhhhuh***uhu-h-h --.hhhh- different times:: - because*[1] *why - .hh*hh -becu:::z (1.8) M: uumm ((2)) M: {lip suck/squeaks twice} ((3))	1- R turns back to WS, adjusting hair, then drops to LB 2- M grabs mouse, R raises to WS grabbing for mouse but aborts when he sees cursor move (1.0) 3- Both watch as M clicks RUN 8-9 times (9.3)	3- CVCK starts to run continuousl y for nine cycles

Segment: AV3p17

The exchange shown in segment FF4p1 presents a typical example of nonverbally-regulated cursor management, as R simply takes control of the shared cursor after the two participants decide on a course of action. Since there was no explicit verbal negotiation to establish R's turn at control, R was clearly relying on nonverbal displays like M's continued gaze at the laboratory manual and the fact that M had not moved his hand towards the mouse to infer that the shared cursor was available.

The value of hand position with respect to the mouse as a resource for regulating access to the shared cursor is emphasized in segment FF4p14, in which participants clearly orient to this nonverbal display to regulate access to the cursor. As the segment begins, R is moving the mouse as the participants try to decide how to get rid of a gauge that they have incorrectly attached to the construction. Rather than verbally requesting control of the shared cursor, M simply places his hand on the mouse; this tacit request for control of the shared cursor is subsequently recognized by R, as he lifts his hand from his mouse to mark the end of his turn at control.

Finally, the movement of the shared cursor in the workspace also constitutes a strong nonverbal resource for regulating access to the shared cursor since, obviously, if the shared cursor is moving, it implies that a partner is currently in control. In segment AV3p17, R can be clearly seen to orient to this nonverbal resource, as she aborts her movement to grab her mouse when she notices the movement of the shared cursor in the workspace.

6.2.1.3 Summary: Resources for Cursor Management

An examination of the cursor management activities engaged in by participants revealed that participants relied on a variety of verbal and nonverbal resources to negotiate mutually-exclusive access to the shared cursor. Two distinct approaches to cursor management were identified: Verbally-regulated cursor management was characterized by explicit verbal negotiation over current or upcoming control of the shared cursor; Nonverbally-regulated cursor management was defined by the absence of such verbal negotiation, with participants relying on nonverbal displays like hand position, direction of gaze, and movement of the shared cursor in the workspace to tacitly regulate access to the cursor.

It is important to emphasize that the distinction drawn between these two approaches to cursor management is not meant to imply that participants relied *exclusively* on one type of communicative display or the other to regulate access to the cursor in a given situation; verbal and nonverbal displays are mutually constitutive and inevitably contribute to participants' interpretation of each other's current beliefs regarding control of the shared cursor. Rather, the distinction between verbally-regulated and nonverbally-regulated cursor management merely provides a framework for characterizing the resources for cursor management available to participants at any given point in the interaction; under verbally-regulated cursor management, participants had access to both verbal and nonverbal displays as evidence of a partner's orientation towards the shared cursor, while only nonverbal evidence was available in situations in which participants used nonverbally-regulated cursor management.

6.2.2 When Did Cursor Turntaking Breakdowns Occur?

Having characterized the kinds of resources that participants relied on to organize their cursor management activities, we are ready to examine the circumstances under which Cursor turntaking breakdown occurred in the audio-only and audio-video environments. As discussed earlier, the goal of this analysis was to expose consistent correlations between Cursor turntaking breakdowns and the communicative resources that participants were relying on to regulate access to the shared cursor at the time the breakdown occurred, thereby implying that participants' access to those resources was somehow constrained by the technologically-mediated environment.

The analysis revealed that Cursor turntaking breakdowns were overwhelmingly associated with nonverbally-regulated cursor management. That is, the Cursor turntaking breakdowns that were documented in audio-only and audio-video interactions were consistently associated with situations in which participants were relying on tacit, nonverbal resources to regulate access to the shared cursor; no Cursor turntaking breakdowns occurred in which participants relied on verbally-regulated cursor management. The following exchanges present several examples of Cursor turntaking breakdown that illustrate the unreliability of tacit, nonverbally-regulated cursor management in technologically-mediated interactions:

•	R: is it going[7] this way:: or is it going this way -- you know? M: mmhmm --- I dunno ((8)) M: .hhhhh-hhhhh [R: what[9] is it+ - what happens [10]down here --- down at this:: [11]-- part here	7- on first “this”, R traces cursor through H1 from right to left, for second “this” motion is vice versa. 8- R glances RS, then back to WS, M sits up and grabs mouse (1.7) 9- cursor jerks as both control it, M gets it up to control panel for “here” 10- R finger points to C1 then drops hand to mouse. 11- R rolls cursor down to C1	
•	((1)) M: I’ll push it ((2)) R: okay::	1- cursor jerks as both try to control (.5) 2- R moves hand off mouse (.5)	

Segment: AV3p12

•	((1)) R: you wanna do tho:se ones? (.5) M: *I think we haveta have* it hooked [2](.5) [3]yaknow ((4)) M: awwwooh+ -wu+ -- let go of it R: [5]ooh -- there	1- M drops mouse to scratch her head, R rolls cursor over to L in pallete (.7) 2- R gazes to LB 3- M grabs mouse and jerks cursor over near H1 4- R raises gaze to WS and cursor wobbles and jerks across construction as both control mouse (2.5) 5- R jerks his hand off mouse	

Segment: AV3p6

          •	R: *youcn go ahead and take* that thing thats-= M: =*okay:: - can I::: --- start from::::* ((4)) M: uuuuhhh [R: *the* middle or whatever (.9) M: uh-- *the center?* R: sure ((5)) M: *oh*hh I don't know what [6]----- this = [R: hey M: = is-, the center?	4- M raises to WS, grabbing mouse, R drops to LB, M slowly rolls cursor down pallete (1.8) 5- R inches cursor back upward towards H (.5) 6- Cursor jerks momentarily and then M centers it on H	

Segment: AO4p3

The exchange shown in segment AV3p12 presents a typical example of Cursor turntaking breakdown related to nonverbally-regulated cursor management. Initially, R has control over the cursor, using it to illustrate the flow of blood through the ventricle (H1). The Cursor turntaking breakdown occurs when M, apparently assuming that R's turn at control had ended during a brief (1.7 second) silence, attempts to take over control of the cursor. Note that, though R appears to momentarily regain control over the cursor, a second Cursor turntaking breakdown occurs subsequently, indicating that the participants remain confused about who controls the cursor. Ultimately, the confusion is resolved is by explicit verbal negotiation of who controls the cursor.

This pattern of breakdown behavior, in which a series of Cursor turntaking breakdowns associated with nonverbally-regulated cursor management is followed by the use of verbally-regulated cursor management to repair the persistent confusion, was very common in technologically-mediated interactions. This observation emphasizes that the presence of Cursor turntaking breakdown is, in itself, a relatively poor resource for repairing the breakdown; even when it is evident to both participants that confusion over who controls the shared cursor exists, nonverbal resources were apparently unreliable for re-establishing mutually exclusive access to the shared cursor.

The exchanges shown in segments AV3p6 and AO4p3 present another common pattern of Cursor turntaking breakdown associated with nonverbally-regulated cursor management. In both of these exchanges, control of the cursor is initially negotiated verbally; Cursor turntaking breakdowns occur when participants attempt to rely on nonverbal resources to tacitly negotiate a subsequent transition in cursor control. For instance, as segment AV3p6 begins, R verbally offers control of the cursor to her partner, opening an explicit negotiation over who should control the cursor. However, when M fails to verbally respond to this overture, R assumes that he retains control of the cursor; Cursor turntaking breakdown results as both participants move the cursor. The exchange shown in segment AO4p3 is similar, with the participants first using verbally-regulated cursor management (i.e. "go ahead and take that thing") to negotiate control over the shared cursor. The Cursor turntaking breakdown occurs when M subsequently usurps control of the cursor as she raises her gaze back to the workspace, apparently not aware that R has been inching it across the workspace.

In each of the exchanges presented above, Cursor turntaking breakdown is clearly associated with nonverbally-regulated cursor management, occurring in situations in which participants relied primarily on nonverbal resources to regulate access to the shared cursor. A vitally important observation is that, in each case, the Cursor turntaking breakdown occurred *in spite of* compelling nonverbal displays of cursor control produced by participants. For example, in segment AV3p12, M attempts to take control of the shared cursor despite the fact that R is still gazing at the workspace with his hand on his mouse, providing strong tacit evidence of his continuing control of the cursor. Similarly, the breakdown in segment AV3p6 occurs as R apparently fails to notice that M has tacitly accepted his preceding verbal offer of control over the mouse by moving her hand to the mouse. Finally, in segment AO4p3, the Cursor turntaking breakdown occurs as M attempts to control the cursor despite the fact that her partner is gazing directly at the workspace, has her hand on the mouse, and is moving the cursor slightly, providing strong nonverbal evidence that she believes to be in control of the shared cursor.

These observations clearly imply that participants in technologically-mediated interaction were insensitive to nonverbal displays of cursor control like direction of gaze, hand motion and position with respect to the mouse and, consequently, were unable to utilize these resources to inform their cursor management activities. As a result, participants were deprived of vital evidence of a partner's current beliefs regarding control of the shared cursor, resulting in a greater likelihood of Cursor turntaking breakdown.

6.2.3 Cursor Management in Copresent Interaction

One possibility that has not been addressed in the preceding discussion is that reliance on nonverbal resources to regulate access to the shared cursor was inherently error-prone, leading to Cursor turntaking breakdown in *any* communication environment. An analysis of participants' cursor management activities in copresent interactions showed that this was emphatically not the case. Despite the fact that copresent participants relied almost exclusively on the tacit, nonverbally-regulated cursor management strategy, Cursor turntaking breakdown was almost non-existent in copresent interactions, with only two breakdowns occurring over the course of all four copresent interactions.

Unlike their counterparts in the technologically-mediated environments, copresent participants displayed an intimate awareness of the nonverbal displays of their partner, clearly relying on such resources to inform their cursor management activities, as illustrated in the following segments:

	M: just like a puzzle ((5)) M: urr[6]!	5- M finishes positioning L2, R checks LB briefly then gazes WS (1.3) 6- M places L2 as R watches	6- flash/connect. construction now complete, except they have all Us and no Vs!
•	(1.0) M: (*cannot scope)* R: kay now lets go up there to run (1.0) R: [1]nn click (1.3) M: *do* wha::t? ---[2]--- you do that - part ((3))	1- M gazes LB with hand on mouse, R jerks hand towards mouse then relaxes as she speaks (1.0) 2- R grabs mouse and rolls towards control panel 3- R centers on RUN and clicks, they watch (3.2)	3- CVCK runs a cycle, arrows in pipes move, etc.

Segment: FF2p8

•	R: how do we get rid of[2] this? ((3)) M: try the[4]::: ((5)) R: tr-hi sump'mm [M: sump'mm (.7) R: hummm[6]mmm ((7))	2- R points and clicks on G2a graph with cursor 3- R rolls cursor back and forth, M glances towards the table (1.7) 4- M grabs his mouse 5- R rolls mouse some more, then pulls back hand and shrugs as he speaks (1.2) 6- M clicks on G2a 7- M hesitates, then drags G2a over to the biowaste (4.2)	2- G2a highlights 7- biowaste highlights as it is contacted.

Segment: FF4p14

In the exchange shown in segment FF2p8, R starts to reach for her mouse but then aborts her hand movement, apparently concluding that M still believes to be in control of the shared cursor. In particular, R's behavior suggests that she was able to utilize nonverbal resources to arrive at two insights that collectively inhibit her from taking control of the mouse:

1. M may still believe she is in control of the shared cursor, since she was the last to use the cursor and displays continuing evidence of her control in that her hand remains on the mouse.

2. M is not able to perceive R's nonverbal request for cursor control (i.e. reaching for the mouse) since she is gazing intently at the laboratory manual at the moment.

The second of these observations emphasizes that regulating access to the shared cursor is a *negotiated* activity, in which each participant monitors his or her partner's communicative displays for evidence of the partner's evolving interpretation of action. Specifically, R is aware not only of M's continued nonverbal display of cursor control, but also of the fact that M was unable to perceive R's own nonverbal request for control of the shared cursor. Only after M returns her gaze to the workspace does R repeat her hand movement to grasp her mouse and take control of the shared cursor, now confident that M is able to perceive her nonverbal demonstration of control.

The exchange shown in segment FF4p14 presents an even more compelling example of how copresent participants are able to utilize nonverbal resources to regulate access to the shared cursor. From a strictly verbal perspective, M's aborted utterance "Try the:::" appears to be a request for R to perform some action. The fact that M's utterance actually represents an implicit request for control of the cursor is only apparent in light of M's movement of her hand to her mouse. The subsequent transfer of control over the shared cursor is then progressively negotiated entirely by nonverbal means, with both participants clearly orienting to each other's nonverbal displays to inform the transaction. Specifically, M does *not* immediately move the cursor after issuing her tacit request for control of the cursor, demonstrating her awareness that R is still using the cursor and is not yet ready to give up control. Only after R tacitly acknowledges M's request for control by removing his hand from his mouse does M actively assume control of the shared cursor. Table 6.1 accentuates the role of nonverbal resources in regulating this tightly choreographed exchange.

Table 6.1: Nonverbal resources used to inform the negotiated transfer of cursor control from one participant to other

Step in negotiation	Nonverbal displays
Initial state: M is aware that R controls the cursor.	Awareness of R's gaze, hand position, and movement of the shared cursor in the workspace.
Request: M tacitly requests control over the shared cursor.	Mutually constitutive significance of M's utterance and movement of hand to mouse.
Postponement: R maintains control in order to finish current action; M is aware of this and waits.	Mutual awareness of R's direction of gaze (on workspace), hand position (on mouse) and motion of cursor in workspace.
Acknowledgment: R finishes acting and relinquishes control, acceding to M's earlier request.	R's shrug and movement of hand from mouse coupled with verbal invitation to take control.
Confirmation: M accepts control of the shared cursor	M's direction of gaze, hand position, and motion of cursor in workspace.

This characterization of the exchange shown in segment FF4p14 clearly emphasizes the intimate awareness that copresent participants have of a partner's nonverbal displays, and how such awareness serves to regulate access to the shared cursor. In particular, it demonstrates how participants rely on the mutual accessibility of such behaviors, not only to infer a partner's beliefs about who currently controls the shared cursor, but also to provide "feedback" during tacit negotiation of cursor control.

In sum, analysis of cursor management in copresent interactions revealed that copresent participants were clearly sensitive to the nonverbal displays of their partners, and were able to consistently use these displays as a reliable basis for negotiating mutually exclusive access to the shared cursor.

6.3 Rationalizing Reference Breakdown

In communicating about their task context, participants had to continually establish the relationship between referential terms that appeared in their utterances, and the objects and spaces within the task context to which those terms referred. Reference breakdowns occurred when this process somehow failed, causing uncertainty as to whether both participants had identified the same object as the referent of an immediately preceding utterance; such confusions were evidenced through explicit verbal repair aimed at clarifying the reference, initiated either by the speaker or the listener.

6.3.1 Resources for Reference Management

An analysis of the reference management activities of participants in all three environments revealed that speakers produced a variety of verbal and nonverbal displays to identify the referents of their utterances. As a framework for analysis, two conversational mechanisms for establishing shared reference were identified: *verbal description* and *deictic gesture.* Both of these mechanisms characterize distinct communicative resources provided by a speaker *in addition* to the reference itself, aimed at somehow constraining the listener's search for the appropriate referent. Speakers using deictic gesture augmented references with nonverbal resources by pointing to referents using either their fingers or the shared cursor. Speakers using the verbal description mechanisms identified referents by describing their physical appearance or spatial location. The following sections discuss each mechanism for reference management in more detail.

6.3.1.1 Deictic Gesture for Reference Management

Deictic gesture was by far the most common mechanism used by speakers to identify the referents of their utterances. Use of the deictic mechanism was characterized by a pointing action made by the speaker using either a finger or the shared cursor[2] that occurred directly before, during, or directly after the speaker produced a reference in an utterance. Several examples of direct reference supported by deictic gesture are shown in the following segments:

•	M: wait --- does [5]it look?- doesn't it look like it has[6] *a direction?* (.4) R: yeah:: it does::	5- M leans forward and finger points to the new V 6- M retracts pointing finger and brings it to bear on LB as he drops his gaze.	

Segment: FF5p2

•	((3)) R: oh rotate[4]--i see [M: .hhh - hehehe ((5)) R: rotate	3- Both chuckle more as R moves cursor around aimlessly and M gazes LB (1.5) 4- R moves cursor up to rotate button, then to L then back to rotate button. 5- M still chuckles lightly, R goes up and clicks the rotate button (1.3).	5- Since no component is hilighted, dialog pops up warning to hilight before rotate

Segment: AO2p3

•	M: *okay::* ((3)) M: now we nee::d a:: 4------ a s::quiggly5 thing:: ((6)) M: oops	3- M snaps to WS grabbing mouse, R pulls hand from mouse (1.1) 4- M rolls cursor up to H in pallete 5- R snaps to LB 6- M clicks on H, then rolls cursor to WS as R returns gaze to WS, then M rolls cursor back towards pallete (2.5)	6- H in pallete hilights

Segment: FF5p2

The exchange shown in segment FF5p2 presents a typical example of how deictic gesture was used to support direct reference. In order to call R's attention to the fact that they are using the wrong component in their construction, M first points to the component icon in the workspace, then to the correct component depicted in the laboratory manual, thereby providing a strong nonverbal resource for locating the appropriate referent[3] of the generic pronoun "it" each time it occurs in M's utterance. The exchange presented in segment AO2p3 demonstrates how the shared cursor was used for deictic gesture as well. In this exchange, the participants are trying to figure out how to rotate components; when R notices the "rotate" button near the top of the workspace, she clarifies the referential significance of her verbal exclamation by bringing the cursor to bear on the icon she is referring to.

In both segments FF5p2 and AO2p3, the status of the speaker's gesture as a deictic resource for establishing shared reference is obvious — the gestures are clearly produced expressly for the purpose of identifying the referents of the speaker's utterance. In task-oriented contexts, however, where participants are busy manipulating various objects in their environment, deictic gesture is frequently conflated with manipulative action. For instance, a movement of the shared cursor across the workspace can serve a distinct deictic purpose, disambiguating a referential utterance made by the participant controlling the shared cursor, while simultaneously playing a role in some ongoing manipulation of the workspace. Similarly, a hand movement that ultimately ends in marking of an answer in the laboratory manual may simultaneously serve to draw attention to the question to which a speaker is referring in an ongoing utterance. In this way, the distinction between deictic gesture and manipulation is blurred in task-oriented interactions, with a tool (i.e. the hand or the shared cursor) serving both as a resource for constructing the referential significance of a speaker's utterance, and as a tool for manipulating the task environment.

An example of this behavior, in which deictic gesture is combined with manipulative action, is shown in segment FF5p2, in which M is trying to decide which component to use next in the process of building the construction depicted in the laboratory manual. M's movement of the cursor to the ventricle (H) as she speaks clearly serves a deictic purpose, establishing which component she is referring to with "squiggly thing." At the same time, the gesture constitutes the first half of the manipulatory action of dragging the a new ventricle into the workspace for use in the construction that the participants are building.

6.3.1.2 Verbal Description as a Resource for Reference Management

Another conversational mechanism commonly employed by speakers to identify the referents of their utterances was the verbal description of referents. Using this mechanism, reference to an object was made by describing its physical characteristics or its spatial location with respect to other objects or entities[4] in the task context, as illustrated in the following examples:

•	((8)) M: *and I know that thing on top is the heart* ((9)) M: cuz you can see it now when you:: - when you hit the[10] run thing:: --- see howit ---[11] expands like that	8- Both marking LB (5.3) 9- M audibly drops pen and raises to WS grabbing mouse (.7) 10- M hits RUN 11- R finishes marking and snaps to WS	10- CVCK runs

Segment: AV2p12

	M: *mmm*[8]*mmm (.5) {ahem}* ((9))	8- R drops to LB 9- Both stare LB, cursor still on gauge icon (1.4)	
•	R: *are the gauges* those little [1]circle thi*ngs?* (1.1) M: mhmmm ((2)) R: *or're -- those where they're sposed to be* (.9)	1- M snaps to WS 2- M drops to LB (.6)	

Segment: AO3p25

	R: *we dunno -- we could be tricked* [M: *well whats yours look like* (.6) R: *mine is a little squa::re (1.3) with like* (.6) *inside the square theres* (1.7) *like two:::* (.7) *like a::, rectangular type*[1] *thing::?* M: *mhmmm -nn then theres sumpin:::* *- some big <u>bub</u>bles, sticking up at the top?=* [R: *two* R: *=at the top and then at the bottom kindof a=* [M: *yea:::h* R: *=little blur::b*	Both participants gaze at their laboratory manuals thoughout the exchange.	

Segment: AO3p5.1

In the exchange shown in segment AV2p12, M uses verbal description to identify the referent in her utterance by spatially locating it as "thing on the top." In segment

AO3p25, the description used by R to identify the referents of her utterance focuses on their physical characteristics, describing them as “those little circles things.” Finally, the exchange shown in segment AO3p5 presents an extended sequence of reference by verbal description as the two participants, who have become uncertain that they have been given identical laboratory manuals, collaboratively describe the cardiovascular construct depicted in the laboratory manual; reference to the various components in the construction is made by describing them in terms of either their physical characteristics or location within the construction.

<u>6.3.1.3 Summary: Resources for Reference Management</u>

An examination of the reference management activities of participants in all three environments revealed two distinct conversational mechanisms that were used by participants to establish shared reference to the entities and objects in the workspace. The most prevalent of the two mechanisms was deictic gesture, in which a speaker nonverbally identifies the referent of a co-occurring utterance by pointing to the object or place in question using either a finger or the shared cursor. A second mechanism was the use of verbal description to somehow characterize the referent’s physical appearance or location. Though speakers tended to rely on one mechanism or the other in a given situation, it was not uncommon for them to make available both resources, as illustrated in the following example.

•	R: right --- ok- what? ((6)) R: wait --- that thing on the botto[7]m? M: it loo[8]ks funny ---- [9]that ones wrong [R: (look at this) ------ yea	6- Both snap to gaze LB, then raise to WS (1.7) 7- R brings cursor down to bottom of construction 8- Cursor jerks about; M is trying to control cursor too 9- M wins cursor and brings it to rest on C1u.	

Segment: AO5p6

In segment AO5p6, R has noticed that the pair has incorrectly placed one of the components in the construction; in referring to the erroneous component, R provides both a verbal description of its location ("that thing on the bottom") and a deictic gesture with the shared cursor.

6.3.2 When Did Reference Breakdowns Occur?

Having characterized the communicative displays used by participants in general to establish shared reference, we are prepared to discuss the relationship between these resources and the Reference breakdowns that occurred in technologically-mediated interactions. Episodes of Reference breakdown that occurred in the audio-only and audio-video interaction were examined, looking for consistent patterns in the communicative resources that participants were relying on to maintain shared reference when breakdowns occurred. This analysis revealed three distinct patterns of Reference breakdown:

1. Reference breakdowns were often associated with the lack of deictic gesture. That is, Reference breakdowns occurred when speakers made totally unsupported pronomial references or relied on verbal description to identify the referents of their utterances.

2. Reference breakdowns occurred when deictic gesture was provided by a speaker, but the listener failed to *perceive* that gesture because he or she was gazing in the wrong direction at the time the gesture occurred.

3. Reference breakdowns occurred when the speaker became uncertain about the adequacy of an immediately preceding reference, due to an apparent inability to monitor a partner's interpretation of that reference.

In the following sections, each of these patterns of Reference breakdown is discussed in more detail.

6.3.2.1 Reference Breakdown in the Absence of Deictic Gesture

Reference breakdowns regularly occurred when speakers relied primarily on verbal resources to make available the referent of their utterances, failing to support their references with deictic gesture. Such breakdowns occurred either because the speaker produced a pronomial reference not supported by *any* additional verbal or nonverbal displays, or because the speaker relied on the verbal description mechanism discussed earlier to identify a referent. The following segments present several examples of this behavior:

•	((1)) M: do we gotta make that smaller? ((2)) R: this? (.5) M: yeah R: I dont know	1- R clicks to disappear the dialog, then precisely aligns V1, then check LB and moves H1 into place. M is gazingWS and checks LB several times (10.9) 2- R hesitates, glances LB, then back to WS where she puts mouse on H1 (2.2)	1- flash/connect H1

Segment: AV5p11

•	((3)) M: we need *that one with the lines in it* ---- [4]for up there dont we? (1.1) R: *which one?* (.6) *oh* it has the li::nes (.7) okay[5] I see ((6)) R: this one? M: yea::h[7] -- I think so (.8) yea::h	3- R moves in a new L, rotates it twice, places L1, moves H1 out of way a little, drags in another U and begins to (erroneously) position it as V1 (27.9). Both R and M gaze WS but check LB several times. 4- R gazes LB 5- R clicks to grab the erroneous U 6- R moves erroneous U up and next to palette by the extra T, then rolls cursor to the V in pallete. (4.3) 7- R clicks V and drags into WS	3- New L appears in WS, flash/connect as L1 placed, new U appears in WS, flash/connect, as it's placed as V1 7- New V appears in WS

Segment: AV5p10

•	R: capillary::: (1.2) (lets se+) umm ((2)) M: hehe --- .hhhh - what -- what was the thing above the bottom thing (1.0) was that just a R: [3]this? M: yea::h (.5)	2- R gazes LB in apparent excitement, M idly moves to close the dialog but misses with the click, R returns to gaze WS while M now gazes LB for (2.0), then both stare WS as R closes dialog, then pulls down and views menus on far right, then moves the extra C in WS over with other extra components (36.0) 3- R moves cursor onto L in palette	

Segment: AV5p14

The Reference breakdowns presented in segment AV5p11 occurs when the speaker produces a pronomial reference (i.e., "that") without providing any additional verbal or nonverbal resources for locating the appropriate referents. In essence, the reference in these examples is anaphoric, with the speaker relying on the tacit context of the interaction to disambiguate his or her reference.

The Reference breakdowns presented in segments AV5p10 and AV5p14 are associated with the failure of the verbal description mechanism to unambiguously identify the referent of a speaker's utterance. In segment AV5p10, for example, the Reference breakdown occurs as M tries to point out that R has used an incorrect component, placing an unvalved vessel (U) where a valved vessel (V) should be; to refer to the erroneous component, M describes its physical characteristics (e.g., "the one with the lines in it"). Similarly, the Reference breakdown in segment AV5p14 arises from the apparent inadequacy of M's description of a referent by its location within the construction (e.g., "the bottom thing").

In sum, Reference breakdowns regularly occurred in technologically-mediated interactions when a speaker failed to provide deictic gesture as a resource for locating the appropriate referent of an utterance, relying instead on verbal description of the referent or the conversational context to identify the referent. Clearly, these referential mechanisms do not provide a reliable basis for negotiating shared reference.

While these observations establish the existence of a consistent pattern of referential trouble, they do not shed any light on *why* this pattern of trouble was so prevalent in technologically-mediated interaction. How can these Reference breakdowns be rationalized in terms of resource constraints imposed in the audio-only and audio-video environments? An answer to this question is indirectly revealed by the observation that Reference breakdowns associated with missing deictic gesture were almost invariably *repaired* using deictic gesture to point out the appropriate referent; this repair behavior is evident in each of the exchanges presented above. This observation clearly implies that deictic gesture is generally a *much stronger* resource for identifying referents than verbal description. The obvious question, then, is why speakers did not *always* support their references with deictic gesture. That is, why would a speaker ever fail to provide a deictic gesture, relying instead on the verbal description of the referent, when verbal description is clearly more prone to Reference breakdown?

A more detailed examination of the situations in which speakers resorted to verbal description of referents revealed a compelling answer to this question: speakers resorted to verbal description *when they did not have access to the shared cursor* at the time they produced an utterance. That is, the shared cursor was under the control of the speaker's partner at the time of the utterance, rendering it unavailable to the speaker as a deictic tool. For example, consider the following exchange:

•	(.4) R: I know ((1)) R: duu:::h[2] M: go up to where it says help R: go duh r::otate ((3)) R: [4]help? (1.1) R: does(n't)--ohh ((5))	1- R drops G up and away from H1 (.9) 2- R rolls mouse up to rotate button 3- R hesitates, looking confused (1.1). 4- M is just dropping gaze to LB, but returns to WS as R speaks 5- R clicks on rotate and both stare at dialog (1.9)	5- dialog pops up saying that gauges dont rotate

Segment: AO2p12

In this exchange, the participants are trying to decide how to attach a gauge. Just as R hits on the (erroneous) idea of rotating the gauge and rolls the cursor to the rotate button to try it, M suggests looking under the HELP menu. The Reference breakdown is evidenced by R's apparent confusion[5] regarding the referent of M's directive to "go up to where it says help." What is important about this exchange is that R is controlling the shared cursor at the time that M uses the verbal description mechanism to identify the referent of her utterance; the shared cursor was not available to M as a deictic tool.

This observation holds for the exchanges shown earlier as well; in each case, the speaker that used the verbal description mechanism to identify a referent did not have access to the shared cursor. For instance, in both segments AV5p10 and AV5p11, R clearly has control of the cursor throughout the exchange, making it unavailable to M as a deictic tool as she produced the problematic reference; in segment AV5p14, R was also controlling the shared cursor at the time M produced the ambiguous reference.

From these observations, we can conclude that speakers may have preferred to provide deictic gesture as a resource for constructing the referential significance of their utterances, but were unable to do so because the shared cursor was being used by their partner at the time. Consequently, speakers were forced to resort to verbal description of referents, frequently resulting in Reference breakdown when such descriptions failed to unambiguously identify referents. In this way, the Reference breakdowns associated with the absence of deictic gesture can be rationalized by the fact that only one cursor was available as a deictic tool in the environment.

In sum, one distinct pattern of Reference breakdown was characterized by the failure of the speaker to provide deictic gesture as a resource for establishing shared reference, relying instead on verbal descriptions of a referent's physical appearance or location. A rationale for this pattern of breakdown was suggested by two observations:

1. In almost all cases, the Referential breakdowns associated with the use of verbal description of referents was eventually repaired through the use of deictic gesture.

2. The shared cursor was invariably unavailable to speakers using verbal description of referents at the time they produced their utterance.

These observations strongly imply that deictic gesture is a much more reliable mechanism for establishing shared reference, but that speakers were forced to resort to verbal description when shared cursor was not available as a deictic tool. In this way, the availability of a single shared cursor is clearly implicated as a causal factor for this pattern of Reference breakdown.

One question raised by this discussion is why this pattern of breakdown, which was related to the unavailability of the shared cursor as a deictic tool, was just as prevalent in audio-video as in audio-only interactions. While the shared cursor is clearly the *only* mutually-available deictic resource in the audio-only condition, it would seem that participants in audio-video interactions could have used their fingers to produce deictic gestures, rather than resorting to verbal description of referents. The evidence presented in the following section provides one possible answer to this quandary, by revealing that participants in audio-video interactions were largely insensitive to finger deixis by their partners, making them even less reliable than verbal description of referents as a resource for establishing shared reference.

6.3.2.2 Reference Breakdown When Deictic Gesture was Available

A second pattern of Reference breakdown was defined by situations in which a speaker provided deictic gesture in support of a reference, but the listener somehow *failed to perceive* that gesture. As a result, the speaker's deictic gesture was essentially unavailable to the listener as a resource for constructing the referential significance of the speaker's utterance, resulting in a greater likelihood of Reference breakdown. This pattern of Reference breakdown occurred both in conjunction with cursor deixis and finger deixis; in both cases the underlying problem was that listeners were apparently unaware that a deictic gesture was being produced by the speaker and, consequently, failed to attend to the gesture.

6.2.2.3 Insensitivity to Finger Deixis

A particularly pervasive pattern of Reference breakdown in technologically-mediated interactions was associated with situations in which speakers used their fingers to deictically identify the referents of an utterance. Listeners appeared to be totally insensitive to such gestures, frequently leading to Reference breakdown. For example, consider the following segments:

•	R: how do you change it ((2)) R: h[3]nhnhnnhnhnh ((giggling)) [M: that arrow right [4]there ((5)) M: click on that arrow R: this one? M: yeah	2- R clicks "beats" box again, then moves cursor to slider and rolls it back and forth over it without clicking (4.3). 3- R scribbles cursor wildly over slider while laughing. 4- M points a finger at WS, nodding. 5- R moves cursor to right sliderbar arrow (.3)	2- beats box highlights when clicked.

Segment: AO2p8

	((5)) M: What [6]about the[7] second thing down	5- R gazes LB while M still gazes WS (2.3) 6- M points to WS 7- R gazes WS and grabs mouse	
•	R: (um that) [1]this one? M: yeah:: ((2))	1- R rolls cursor down to V in pallete 2- R does a "describe" on V, they read it (5.0) then R nods and gazes LB (1.0)	2- Describe for V pops up. V stays hilighted.

Segment: AV5p3

•	R: which one is the gauge michelle::, is it+? ((8)) R: *lets see its showing*[9] *these little --- thing::s* (.8) M: *what* ((10))	8- Both gaze to LB (1.1) 9- R puts finger on figure 2 in LB 10- Both stare LB, then R snaps WS to speak (1.4)	

Segment: AV3p21

The exchange shown in segment AO2p8 occurs as the two participants are trying to change the number of beats (i.e. cycles) that the CVCK simulator will run. As M calls her partner's attention to "that arrow right there," she uses her finger to point to her workspace screen to indicate which arrow she is referring to. The Reference breakdown is revealed as R initiates a repair sequence, using the shared cursor to suggest a possible referent of M's utterance. Since this exchange occurs in an audio-only interaction, it is clear that M's deictic finger gesture was fundamentally unavailable to R, since participants have no visual access to each other. However, the latter two segments presented above emphasize that similar breakdowns occurred in the audio-video condition as well. For instance, in segment AV5p3, M uses finger deixis[6] to identify her referent, as she suggests the next component to install in the construction the pair are piecing together. Note that R never directs her gaze at the remote monitor, gazing instead at the laboratory manual and the workspace as M gestures. That is, R appears to be totally *unaware* of the deictic gesture that M is making available. A similar pattern of behavior is evident in segment AV3p21, in which R points to the laboratory manual to identify the referent of the utterance "... it's showing these little things" while M gazes steadfastly at her own laboratory manual, again clearly unaware of the deictic gesture available in the remote video image.

In sum, Reference breakdowns frequently occurred in situations in which speakers used their fingers to deictically identify referents. Even though participants in the audio-video condition had visual access each other's finger pointing via the remote video image, they were profoundly insensitive to these nonverbal resources, almost universally failing to perceive such deictic gestures. Specifically, they seemed generally unaware that a partner was making a deictic finger gesture available in the remote video image, and consistently failed to turn to gaze at the remote image at the crucial moment to perceive the gesture. The following segment provides a final emphatic example of this insensitivity to deictic gesture in the remote video image:

•	(.7) R: yeah M: [8]isn't that sposed to be:: --- doesn't that connect those corner pieces t+ [[R: *naw but* *I'm*[9]- **I'm talkin about uhhh --- right here**[10] [M: [11]ohhh! ohhh ohhh -- I see[12] ---- you mea::n like <u>this</u> R: yeah	8- R drops to LB 9- M snaps to LB; as R turns to RS, he grabs his LB and holds it up in front of the RS. 10- R moves his gaze from RS to LB as his finger points to L2/L3 in fig1., M is gazing WS. 11- M snaps to WS and jerks mouse down towards L3 12- R snaps back to WS as M grabs L3 and drags it adjacent to L2	

Segment: AV2p4

In segment AV2p4, R makes a concerted effort to make his deictic gesture available to M, holding his manual up in front of the remote image as he points to the laboratory manual. Unfortunately, M remains totally unaware of R's efforts, gazing steadfastly at the laboratory manual. This exchange clearly demonstrates the profound insensitivity of to deictic gestures available in the remote video image, even when such gestures were extremely overt.

6.2.2.4 Insensitivity to Cursor Deixis

Reference breakdowns also frequently occurred in technologically-mediated interactions when speakers provided deictic gestures using the shared cursor. Again, the underlying problem appeared to be that listeners failed to perceive the deictic gesture in question because they were unaware that the gesture was occurring. For example, consider the following exchanges:

	R: did you see it? M: yeah::: (.6) so *I I would say those are open* [R: whats [7]the step thing	7- R rolls cursor to STEP	
•	(.5) M: *huh::?* R: whats --[1]-- step (.7) M: *step?* R: up here[2] at the top ((3)) M: ohhhh, at the top? (.4) R: yeah::, see? [M: ohhh ((4)) R: where the cursor is? M: mhmm	1- M hunches forward and examines LB intently 2- R is wiggling cursor over STEP 3- M still gazes LB, then raises to WS to speak (1.6) 4- R wiggles cursor again (.8)	

Segment: AO4p12

•	M: uuuuum ((5)) M: *jeez* ((6)) R: what is this (u+)deal:: [7]all about (.9) M: what ((8)) R: this thing right her::e ((9))	5- M rolls cursor around pallete, then drops to LB to speak (1.5) 6- M gazes LB, R gazes WS then glances LB then rolls mouse to T (1.2) 7- M raises gaze to WS, moving mouse slightly, cursor jerks off T 8- R recenters cursor on T (.8) 9- first R then M drop to LB (2.5)	

Segment: AV3p23

In both segments AO4p12 and AV3p23, Reference breakdown occurs when the listener fails to perceive the cursor deixis made available by the speaker because he or she is not attending to the shared workspace when the gesture is produced. In segment AO4p12, R identifies the referent of her utterance "what's the step thing" by pointing to the STEP button in the workspace; M fails to perceive this gesture because she is gazing at the laboratory manual at the time that it occurs, apparently under the impression that R is referring to something in the manual. The fact that M continues to gaze at the laboratory manual until R explicitly calls her attention to the shared cursor emphasizes M's total lack of awareness of the deictic gestures produced by R.

A similar lack of awareness of cursor deixis is evidenced in segment AV3p23. In this exchange, the pair is in a quandary about which component is the "gauge" that the laboratory manual is asking them to attach to their construction. As R verbally calls attention to a new possibility (i.e. "what's this deal all about"), he rolls the cursor to the component in question. Unfortunately, M is gazing at the laboratory manual at the time, raising her gaze to the laboratory only after the deictic gesture has been completed. To make matters worse, M slightly moves her mouse as she raises her gaze to the workspace, bumping the cursor off of the component being referred to. This behavior clearly implies that M was unaware that R had taken control of the cursor to deictically identify the referent of his subsequent utterance.

The exchanges presented in segments AV3p23 and AO4p12 demonstrate that participants in technologically-mediated interactions were often insensitive to a speaker's use of the shared cursor to produce deictic gesture and, consequently, were not able to reliably attend to such gestures. While participants had no trouble perceiving and interpreting such gestures when they happened to be gazing at the workspace at the time the gestures occurred, Reference breakdowns frequently occurred when a listener *was not* gazing at the workspace at the appropriate moment. Clearly, the problem of realizing that a speaker is using the shared cursor to gesture deictically is essentially a special case of the more general problem of recognizing that a partner is currently in control of the shared cursor. Accordingly, Reference breakdowns related to the failure to perceive cursor deixis can be rationalized in the same way as Cursor turntaking breakdowns, namely, by participants' insensitivity to their partner's hand motions and position with respect to the mouse, and his or her direction of gaze.

In sum, a pervasive pattern of Reference breakdown was characterized by participants' failure to perceive the deictic gestures produced by a speaker. Two distinct variations of this behavior were noted:

1. Participants were almost entirely insensitive to deictic gestures produced when a speaker used his or her finger to point out a referent. While it is obvious why such gestures were not accessible in audio-only interaction, participants in audio-video interactions also displayed a profound insensitivity to such deictic gestures, even though they were available in the remote video image.

2. Reference breakdown also occurred when a speaker used the shared cursor to deictically identify a referent. Breakdowns occurred when listeners were gazing

elsewhere at the time, demonstrating a lack of awareness of the speaker's point of attention and use of the shared cursor.

In sum, whether deictic gestures were produced using a finger or the shared cursor, Reference breakdowns related to the failure to perceive such gestures can be rationalized by participants' insensitivity to the nonverbal resources made available by a partner. Specifically, the failure to perceive finger deixis suggests an insensitivity to the deictic hand motions of a partner; the failure to perceive cursor deixis demonstrates an insensitivity to a partner's point of attention and hand motions with respect to the mouse.

6.2.2.5 Reference Breakdown Related to Speaker Uncertainty

Both of the patterns of Reference breakdown discussed so far have focused narrowly on the verbal and nonverbal displays provided by speakers as resources for identifying the referents of their utterances. In contrast, the third pattern of Reference breakdown exposed in the analysis was related to the inability of the speaker to access the nonverbal displays of the listener. Specifically, Reference breakdowns were defined by situations in which the speaker spontaneously initiated a repair sequence aimed at verifying that an immediately preceding reference had been understood by the listener. The following exchanges illustrate this pattern of Reference breakdown:

•	((1)) R: <u>ro</u>ta2te (.8) R: right here ((3)) M: okay (.4) R: se**e that?** M: 4yeah	1- M drags Lx to top right in WS, R stares WS (.6) 2- R finger points and holds on rotate icon in his WS 3- R drops hand, M drops Lx, and jerks cursor towards "rotate" (.5) 4- R gazes back to WS	

Segment: AV2p3

•	R: *okay:::- r<u>un</u> the simulation-* 2*carefully observing what happens on the screen* ((3)) M: *(sometimes nnn nn necessary is) ---- okay* R: so 4(it must be) up here somewhere? ((5)) M: *hmmm?* ((6)) R: wait ---- (mm-ah) 7under run? ---- dyou see it? (.5) M: mmhm! R: o:*kay::*	2- M gazes LB 3- R pauses then raises to WS (2.0) 4- R slowly rolls cursor to control panel 5- M glances WS and back to LB(.8) 6- M snaps to WS (1.1) 7- R waggles cursor back and forth near RUN	

Segment: AO4p8

In segment AV2p3, the participants are trying to decide how to rotate an elbow (Lx) so that it will fit into their construction. As R drags the component across the workspace, M calls attention to the ROTATE button, pointing to it with her finger. The Reference breakdown is revealed when M initiates a repair, asking R to verify that she has located the referent. A similar situation is presented in segment AO4p8, except that in this case the speaker (R) becomes uncertain that a reference supported by cursor deixis has been correctly interpreted and, again, asks her partner to confirm that she has located the appropriate referent.

In both segments AV2p3 and AO4p8, it is apparent that the referential confusion was not related to the communicative resources made available by the speaker, since it is the speaker who displays uncertainty over whether shared reference has been established. More strongly, no consistent correlation was found between this pattern of Reference breakdown and the communicative resources made available by the speaker; breakdowns were common regardless of whether the speaker used deictic gesture or verbal description to identify referents. Rather, it appears that this pattern of Reference breakdown was related to the inability of the speaker to tacitly determine whether an immediately preceding reference had been successfully interpreted by a listener. This observation emphasizes that establishing shared reference is a *negotiated* process, in which the speaker and listener collaborate to maintain shared reference. In particular, it is not the case that a speaker makes available certain verbal and nonverbal resources for interpreting his or her references, and then simply assumes that those resources were adequate to establish shared reference. Rather, speakers continuously monitor the displays of listeners for evidence that references have been correctly interpreted; such evidentiary displays might include verbal confirmations that the listener has located the appropriate reference (e.g. “okay so where do you want to put [the referent]”) and nonverbal displays like directing gaze at the referent, or manipulating it in some way.

In light of these observations, Reference breakdowns related to speaker uncertainty over the adequacy of immediately preceding references can be rationalized by speakers’ apparent inability to monitor a partner’s interpretation of those references. In particular, the fact that Reference breakdowns of this sort invariably occurred in the absence of strong verbal feedback from a listener suggests that speakers were unable to reliably access nonverbal resources made available by listeners (e.g. direction of gaze), causing speakers to become uncertain whether a reference had been correctly interpreted.

6.3.3 Reference Management in Copresent Interaction

As discussed in the preceding sections, many of the Reference breakdowns observed in interactions that took place in audio-only and audio-video interactions can be rationalized by the availability of a single shared cursor and by the overall insensitivity of distributed participants to the nonverbal displays (e.g. finger deixis, hand position, and direction of gaze) of their partners. An examination of reference management in copresent interactions revealed that copresent participants were not subject to the same constraints. That is, copresent participants were able to non-problematically gesture deictically using both the shared cursor and finger deixis, and displayed an intimate sensitivity to each other's nonverbal displays, clearly using these resources to inform their reference management activities. The following exchanges provide several examples to support these assertions:

	R: to get inta ((inaudible whisper)) first select object, and then pull it ---- then pull dow::n - the help menu ((2)) R: should I --[3]-- click at one of these? [M: no click on[4] this once and then[5] (do it) ((6))	2- R clicks to close dialog (1.0) 3- R rolls cursor over to left pallete of components. 4- M leans points to the gauge icon in WS, R rolls mouse over to gauge icon. 5- M now points to HELP menubar item, then drops hand. 6- R drags in new gauge (.9)	2- dialog goes away 6- A new G appears in the WS. Call it Gb.

Segment: FF2p18

	((5)) M: so this'll need to rota::::te - like-[6] tha:::t ((7)) M: *uum* ((8)) R: ohh I'm sorry ((9))	5- M drops L near V2 (1.2) 6- M has rolled to "rotate" and now clicks 7- Both hesitate at WS, then drop to LB in perfect synch (1.4) 8- Both raise to WS in synch (1.4) 9- M clicks "rotate" again and they stare WS (1.8)	6- L rotates once 9- L rotates again

Segment: FF5p3

	M: got a broken joint? ((8)) R: (somethin) ((9))	8- M gazes LB and flips to read next page again (.8) 9- M reads LB as R realigns L2, then realigns L3 again (10.1)	9- L2 and L3 hilights when moved; L3 hilighted when done

	R: there --- try that out ((1)) R: check it out ((2)) M: yea:::h	1- R clicks WS, then rolls cursor up to RUN in control panel (1.2) 2- M snaps to WS, R clicks RUN (.7)	1- L3 unhilights 2- CVCK starts running

Segment: FF4p7

The exchange presented in segment FF2p18 demonstrates how copresent participants were able to effectively use both cursor and finger deixis to support direct reference. In particular, copresent reference management did not appear to be at all constrained by the availability of a single shared cursor, since (in contrast to audio-video interactions) finger deixis provided an extremely reliable alternative deictic tool that speakers could use when the shared cursor was under the control of the partner. For instance, in segment FF2p18, M is able to deictically identify the two different referents alluded to in her utterance “no click on this once and the do it,” despite the fact that R was in control of the shared cursor at the time of M’s utterance. This provides a powerful counterpoint to technologically-mediated interactions, in which participants often resorted to the verbal description of referents in this situation.

The exchange is segment FF5p3 illustrates how sensitive copresent participants were to each other’s moment-by-moment direction of gaze, able to use this nonverbal resource to maintain an awareness of what their partner was gazing at throughout the interaction. Over the course of the exchange, participants move their gaze from the workspace to the laboratory manual and back again, maintaining a close synchrony in their point of attention. As a result, participants were always aware of what a partner was attending to and, consequently, rarely failed to perceive a partner’s deictic gestures.

Finally, segment FF4p7 illustrates that the awareness of a partner’s point of attention is vital, not only for listeners, *but for speakers as well*. Speakers were able to use this resource to monitor a listener’s interpretation of referential utterances and, in particular, to infer that a listener had *not* perceived a deictic gesture produced by the speaker. As the exchange begins, the participants have just decided that they need to readjust the position of several components before going on to the next step. As R tends to these adjustments, M directs his gaze towards the laboratory manual, apparently reading ahead to see what's next. When R finishes repairing the construction, he accompanies his verbal announcement (i.e. "there, try that out") by rolling the cursor up to the run button. However, R then hesitates and does not proceed to immediately run the simulator, apparently aware that M has not directed his gaze to the workspace. Only after M eventually does direct his gaze to the workspace does R go on to run the simulation again. This exchange emphasizes the keen awareness that copresent participants had of their partner's direction of gaze, using this resource to determine whether deictic gestures had been perceived by a listener and, more generally, to monitor the listener's interpretation of references.

6.4 Rationalizing Topic Breakdown

A fundamental requirement for coherent interaction is that participants must somehow maintain a shared sense of what it is that they are talking about at each moment during the interaction. Topic breakdowns were defined by the failure of this organizational process, resulting in a situation in which one participant believed that the focus of the collaborative interaction had shifted to new topic, while his or her partner still believed the interaction to be focused on the previous topic. Such confusions were evidenced either by explicit repair sequences, in which the issue of current topic was explicitly raised and collaboratively resolved by participants, or by certain verbal and nonverbal evidence clearly indicating that participants had divergent conceptions of the current topic of conversation.

6.4.1 Resources for Topic Management

As discussed in Chapter IV, the overall topic structure of the task-oriented interactions examined in this study is established by the laboratory manual, which gives the sequence of instructions to follow and questions to answer in order to accomplish the collaborative task. These instructions and questions define the primary[7] topics addressed by participants over the course of their interaction.

An examination of interactions in all three communication environments revealed that participants relied on a variety of verbal and nonverbal resources to make available their current topical orientation and to organize transitions from one topic to the next. Nonverbal resources included pointing to the laboratory manual, indexing the laboratory manual with a finger, marking answers, and shifting gaze from the workspace to the laboratory manual. Verbal contributions were used to explicitly negotiate topic transitions, as well to implicitly make available participants' topical orientations. The following sections describe these two classes of topic management behaviors in more detail.

6.4.1.1 Verbal Resources for Topic Management

Participants relied on several verbal displays to make available their conception of current topic and to negotiate transitions from one topic to the next; three distinct mechansisms were identified:

1. Explicit negotiation. Topic transition was accomplished through explicit discussion and agreement to move on to a new topic.

2. Implicit topic introduction. Topic closure and transition to a new topic was implied by the verbal introduction of a new topic.

3. Narration. Participants read aloud to make available their current topical orientation and, in particular, their progress through a reading task.

Each of these verbal mechanisms for topic management are discussed in more detail below, beginning with the following segments, which present examples of explicit negotiation of transitions between topics:

•	((1)) M: *(as preshinnnn-na aahht -- unnnnnn) -- this looks like a really ([2].9) (tsarrible tortuous) test* ((3)) M: *okay::* ((4)) M: *lets go to the next page* (.6) R: unkay	1- both stare WS, then (2.8) M drops to LB using pen as indexer (.9) 2- R drops to LB with pen ready 3- Both make a check in LB (.8) 4- M makes a second check in LB, while R just gazes LB, then both sit up as M speaks (.9)	

Segment: AV3p43

	M: *thats good enough ----- isn't it?* [R: *okay* R: *mhmm* ((5)) M: *nkay*	5- R finishes writing and stares LB (.5)	
•	((1)) R: *nnkay (.5) flip the [2]page* M: *un:kay* [R: *{ahem}*	1- M still writes as R checks her answer, glances WS, then LB and speaks (2.0) 2- R audibly flips the page, M still writing	

Segment: AO3p22

In segments AV3p43 and AO3p22 the closure of the current topic and the shift the next topic are negotiated by raising the issue of "are you ready to go on?" as an explicit digressionary topic in the conversation. For instance, in segment AV3p43, M's utterance "lets go on to the next page" clearly signals her intent to move to the next topic, while at the same time providing R with the opportunity to either accept or reject this shift of topical focus; the suggestion to "flip the page" produced by R in segment AO3p22 performs a similar function.

While the very explicit approach to topic management illustrated in these exchanges provides for very strong topic transitions, it is also relatively cumbersome — the issue of topic must be overtly raised and addressed to negotiate each shift of topical focus. Accordingly, participants relied only infrequently on this topic management mechanism. A much more common verbal mechanism for negotiating topic transition was to simply introduce a new topic, as illustrated in the following exchanges:

•	M: *we wanna pressure ga::ge at the heart at point ay* (.7) R: kay M: *and we want flow gauges --- [0]at bee* and cee ((1)) M: so R: so[2] how do I get that thing up (1.4) M: you doubleclick on it ((3)) M: there you go (.9) generic ga:::ge:: (.5)	0- R clicks G1 1- R rolls cursor onto blank WS (1.9) 2- R clicks in WS then clicks on G1 3- R clicks, then double-clicks G1, M speaks just as dialog pops up (1.9) 4- R clicks "OK"	0- G1 hilights 2- G1 unhighlights and hilights 3- G1 unhilight, then "generic gauge" dialog pops up. 4- dialog goes away

Segment: AV2p15

•	((0)) R: T[1]a:::da:: M: ok*a:::y* ((2)) R: *Ru*n the simulation-do you wanna do that?	0- R grabs U then (1.5) M gazes LB (2.4) while R positions U, then both gaze WS (2.2), then R checks LB (.7), then finishes positioning U as U2 (2.3) while M taps her pencil audibly against her thumb. 1- R drops U into place as U2 2- Both gaze LB (6.6)	1- flash/connect U2. Construction is complete.

Segment: AO2p6

The exchange presented in segment AV2p15 illustrates a ubiquitous verbal mechanism[8] for signaling a topic transition, namely, by asking a question. Participants begin by finishing an organizational discussion of what they should do next; R then introduces the next topic by simply asking “so how do I get that thing up”. A verbal topic transition mechanism unique to task-oriented interactions (Fox, 1993) is shown in Segment AO2p6. After participants negotiate a closing (Beach, 1993) to the previous topic (i.e. “Tada” — “okay”), a new topic is introduced by implicitly referring to the topical framework established by the task-solution process, as R reads the next instruction from the laboratory manual.

Up to this point, discussion has focused on the verbal mechanisms used by participants to negotiate *transitions* from one topic to the next. An equally important topic management activity is displaying progress *within* a given topic. In general, this is not problematic since, during periods of active collaboration, participants’ verbal discussion directly embodies and reflects progress through the current topic. One situation in which this is not the case, however, is when participants are busy reading the laboratory manual. In order to transform reading from an essentially private activity into one that is mutually available, participants often relied on narration to display their progress, as illustrated in the following segment:

	((3)) M: *(sometimes) its hard to tell whats going o::n -- .hhh just by looking at the running simulation since everything is happening so fast sss especially true when you are trying to compare certain -- .hhh (nannna nannna dathaa fosommss disullul)--.hhh (gauges*[4] *to measure and record blood flow or pressure at various places in the cons) ----* [5]*Modify the system you originally constructed- .hhh by attaching gauges at the places marked in figure two*	3- R marks LB, M audibly turns page (4.4) 4- R pauses marking and gazes WS, M is now tracing sentences with pen as indexer as she reads 5- M glances LB, then WS again, pen still ready, frowns and audibly turns the page. As he does so, he shoots a glance at RS, then gazes LB	

Segment: AV3p19

The exchange shown in segment AV3p19 begins just as participants move on to a new topic. After turning the page, M reads the next question aloud, providing her partner with a strong verbal indication of her topical orientation as well as indicating her progress through the reading task. Note that M's narration periodically deteriorates into an incomprehensible whisper, emphasizing the fact that the primary purpose of such narration is to display topical orientation, rather than to convey information in a linguistic sense.

6.4.1.2 Nonverbal Resources for Topic Management

Nonverbal displays of topical orientation were generally shaped by and oriented towards the overall topic structure imposed by the laboratory manual, providing tacit evidence of which of the topics defined by the laboratory a participant was currently working on. Nonverbal displays can be further broken down into two distinct categories:

1. Manipulation of the workspace. Because the task that participants were engaged in involved the manipulation of the CVCK simulator, the manipulative actions of a participant using the shared cursor were an obvious resource for inferring which of the topics defined by the laboratory manual a participant was working on.

2. Actions directed at the laboratory manual. Since the laboratory manual embodied the overall topic structure for the interaction, nonverbal behaviors involving or directed at the laboratory manual represent strong resources for inferring a participants topical orientation.

Representative examples of nonverbal behaviors falling into each of these categories are presented in the upcoming discussion.

One way in which a participant's topical orientation is tacitly made available is by the manipulations performed by that participant on the CVCK simulator in the electronic workspace. For example, consider the following segments.

•	((3)) R: cl:::ick^4! (.4) ah-hah (1.0) o:::okay-your turn M: awright ((5))	3- R chuckles while finishing positioning of L1 (3.5) 4- Releases L1, releases mouse and leans back 5- Both gaze WS as M drags out another L and drops it to right of construction (5.9)	4 - flash/connec t. . l1-v1-h-v2 in place

Segment: AO2p4

•	R: how do you get rid of (this) R: =awwI'll put[3] it down here somewhere ((4)) R: OH bio-<u>waste</u> ---- n<u>iii</u>cc:::e [M: (biowaste) ((5)) M: oops ahhehe-h-h-h [R: heheheh-h ---.hhhh ((6)) M: you gotta (just throw this thing *in there::*) [R: th+ ---- thatd be ----- thatd be like [7]a	3- R drags Ua down to lower left 4- M leans forward to jab finger at biowaste on screen (.6) 5- R drops Ua in biowaste (2.0) 6- R drags out a new V and positions and places it, M glances LB then speaks (2.8) 7- M raises to WS, R clicks V1	3- biowaste flashes as Ua momentarily touches it, and then comes to rest slightly to the side. 5- biowaste flashes and Ua disappears 6- new V appears in WS and is placed as V1, flash/connec t. 7- V1 unhighlights

Segment: FF4p5

In segment AO2p4, a closing to the previous topic is verbally negotiated by participants in the paired utterances "okay, your turn" and "awright" (Beach, 1993) . The new topic of "positioning the elbow as the next component" is subsequently introduced *implicitly* , as M uses the cursor to drag the new component into the workspace. Segment FF4p5 presents a topic transition in which both the closure of the previous topic and the introduction of a new topic are tacitly accomplished by nonverbal action. The current topic, which was introduced earlier by R's asking "how do you get rid of this", is tacitly closed as R disposes of the erroneous component by dragging it to the BIOWASTE icon; the new topic of discussion is then implicitly introduced as R continues on with the construction process, dragging another valved vessel (V) into the workspace.

A second class of nonverbal displays that were used by participants to make available their topical orientations involved actions related to the laboratory manual. For example, consider the following segments:

	(.5) M: *this*[3] -- *this ones true::* (.6) *this one says- there are times when there is: -- flow towards:- the heart at point bee* ([4].4) *(butnn) this one says there are times when therer:: ---- times when there is flow: away* (.4) R: *uhuh*	3- M is pointing to LB with pen, moving it as he speaks, then tracing sentence while reading 4- M jumps indexing pen back towards left margin of LB	

Segment: FF5p16

	(1.1) R: *what direction is the blood* flow --- well ((4)) R: clock*wise* ((5)) M: *.hhhhh -- when blood flows through a valve, is it* [6]*open::* or clo::sed	4- R gazes WS and whirls several rapid clockwise circles above WS with finger (.9) 5- Both gaze LB and M marks LB (3.0)	

Segment: FF4p8

	(.5) R: *they do::nt they=* M: *=so it doesnt flow backwards, so:::* (.6) *right* [R: (*ss*+) ((6)) M: *aw::ri::ght* ((7))	6- Both stare LB as M marks LB, then (5.4) R raises to WS and M marks then (6.6) R drops back to LB as M finishes marking (2.3) 7- M turns the page and places indexing finger on next as he begins to speak (3.5)	
	M: *.hhhhh (smuuuuuuudge* [1]*spudgelookinyohhhh sensor dench tshhhhhhhhhhhhhhhhhhhhhhhhhh*) ((2))	1- M traces sentences in LB as he blurs/mumbles reading 2- Both stare LB in silence, M still tracing sentences (4.9)	

Segment: FF5p9

The exchanges shown in the above segments illustrate several nonverbal mechanisms used to display topical orientation. In the exchange shown in segment

FF5p16, for instance, M uses her pen to index the laboratory manual, providing strong evidence of the topic she is working on as she moves quickly through several topics, proposing answers to a series of questions posed in the laboratory manual. The exchange shown in segment FF4p8 demonstrates that marking answers in the laboratory manual was also powerful resource for topic management, tacitly indicating that a participant considered a topic to be closed. After R introduces a new topic by reading a question from the laboratory manual, M's acceptance of the answer subsequently proposed by R (i.e. "clockwise") is tacitly signaled as M directs his gaze at the laboratory manual and marks an answer to the question; a new topic is then verbally introduced. In segment FF5p9, participants produce a variety of nonverbal displays to negotiate a topic transition. As in segment FF4p8, participants' mutual awareness of answer marking establishes a strong sense of topic closure; transition to the next topic is reified as M turns to a new page. Finally, the next topic is tacitly introduced by M by placing an indexing finger in the laboratory manual.

6.4.1.3 Summary: Resources for Topic Management

Analysis of the topic management activities engaged in by participants interacting in all three communication environments revealed that participants relied on a variety of verbal and nonverbal resources to maintain a shared topical focus. Verbal resources included explicit negotiation of topic closure and next topic, and the posing of questions or reading instructions from the laboratory manual to introduce new topics. Nonverbal displays included manipulation of the CVCK simulation, and various behaviors directed at the laboratory manual, like indexing questions with a finger, page-turning and marking answers. Finally, a participant's direction of gaze was identified as an important resource for inferring his or her current topical focus.

6.4.2 When Did Topic Breakdowns Occur?

The analysis of Topic breakdowns that occurred in technologically-mediated interactions revealed that breakdowns were consistently associated with situations in which verbal evidence of participants' topical orientation was weak or missing. That is, Topic breakdown regularly occurred when participants relied primarily on nonverbal resources like direction of gaze, answer marking, page turning, and manipulation of the CVCK simulator, failing to support such displays with strong verbal evidence of their topical orientations. The exchanges in the following segments provide several examples of Topic breakdown that demonstrate this insensitivity to a partner's nonverbal displays of topical orientation.

	M: *I think its when::* [7]*pressure in the heart decreases* (1.5) M: *should we just check it and move on?*	7- R clicks RUN	1- CVCK starts to run a cycle in slow motion (because they reduced speed earlier)
•	R: yea*sure-uh-h-* M: *uhhho*[1]*kay* ((2)) R: *.huhhh-hhh-.hh* R: *this is*[3] *(a)* [M: *wh+ - when does blood flow towards the heart take place at point cee* R: *ohhh* [4]*woops* (1.1)	1- M marks LB 2- R marks WRONG (last question) answer and goes to turn page (.6) 3- R audibly turns to next page 4- R flips page back	

Segment: AO3p44

In the exchange shown in segment AO3p44, the participants have just finished discussing the second-to-last question on page 3 of the laboratory manual[9], using the explicit verbal negotiation (i.e. "should we check it and move on") to negotiate the closure of the current topic. However, participants' topical orientation diverges when R accidentally marks the answer to the *last* question and then turns the page to go on to the last page of the manual; the Topic breakdown is evidenced by R's utterance of "ooh woops" as the breakdown becomes apparent from R's reading of the next question. What is important about this exchange is that participants are clearly unaware of each other's page-turning behaviors — M is not aware that R has turned to the next page while R is not aware that M has not done so. That is, participants fail to utilize this nonverbal resource to maintain shared topical orientations.

•	M: I dont thinks----*I don't* think so (.8) R: ((affecting sassy)) ok fi::::ne ((1)) R: its always2 flowing towards the heart--it comes out of the heart and goes towards the heart (.4) so its always flowing towards the heheheeart -3-- whether your at point C or not M: *this is true* [R: but I know what it means (.8) okay--4-the flow graph *for C* ((5)) R: it flows when the pressure is high ((6)) M: *where are you at (.4) your not on part three yet* 7*(.4) did I miss something?* R: *No- at the bottom of part three* M: *okay* (1.4) R: *we missed (.4) we didn't do that part (.6) dyou see it?* (1.2) M: *say it again* R: ((clears throat)) *hnhnhnh ---- part three* (1.5) R: *the:::res (.9) ummm (1.7) at the bottom of the page---page three*	1- M marks an answer while R looks at LB then WS (3.6) 2- R is gestureing as she speaks, vaguely shaping the in-out flows. 3- M gazes at LB and throws her hands and body back and forth in gestures too. 4- M audibly turns page to next section 5- M finishes turning page, R stares LB, then WS, gestures vagulely at screen with pen (3.9) 6- R drops gaze to LB (1.3) 7- M pages back to look at previous page	

	M: *uh-huh* R: *does it say when does blood flow towards the heart take place at point C* M: *an[2]d I thought we said never* ((1)) R: oo:::::::::::hhh (1.9) R: ye::s I guess you're right[3] M: *Okay* ((4)) M: *fl::::::lip the page* ((5))	1- M is nodding as she speaks 2- R pauses, then looks up to stare WS (2.8) 3- R gazes LB and marks answer 4- M to next page again (1.3), R also flips to next page but still staring WS 5- R flips *back* to page 3 again! (1.4)	

Segment: AO2p21

The Topic breakdown presented in segment AO2p21 occurs under similar circumstances. In this case, however, recognizing and resolving the topical confusion requires a substantial repair effort. As the exchange begins, the participants are discussing the answer to one of the questions[10] posed in the laboratory manual. As R summarizes the flow behavior that the participants have just collaboratively discovered in the preceding discussion, M has already turned to gaze at the laboratory manual and mark an answer, and eventually turns to the next page, nonverbally displaying her shift in topic. The Topic breakdown occurs when R continues discussion of the previous topic (i.e. "okay -- the flow graph for C"), revealing that participants' conception of current topic has diverged. Shared topical orientation is only re-established after a lengthy repair sequence. Importantly, the Topic breakdown presented in this exchange is, again, associated with a lack of strong verbal displays of topical orientation. In particular, M's transition to the next topic was evident only through her nonverbal behaviors — redirection of her gaze to the laboratory manual, marking an answer, and turning to the next page.

The following segment gives another example of Topic breakdown that occurred in a situation in which a shift in topic was evidenced primarily by nonverbal behavior:

•	R: *is there* flow past point [1]bee::? (1.2) R: there is, [2]isn't there? ((3)) M: yeah ((4)) R: (*noops*) ((mumbled)) ((5)) R: (so) [M: *I just checked* [6]*the second one* **and the last one** [R: *yea::::h (tha wha I was un)* ((7)) R: *hm*	1- M gazes WS 2- R clicks run again, M glances LB 3- R gazes WS, M glances WS to LB several times as CVCK runs (3.3) 4- R grabs pen and marks LB, M stares WS for several seconds, then also grabs pen and marks LB several times (5.5) 5- Both gazing LB until (9.3) R grabs mouse and glances RS (.3) [M is still gazing LB] so R gazes WS and (1.7) clicks RUN, M gazes WS and both watch it run, then M looks to LB and speaks (5.6) 6- R gazes LB 7- R marks LB, then quickly scans other questions using pen as pointer, trying to catch up, when she gets to last one, she gazes WS (.5) then LB. M is gazing LB, occassionally glancing WS (14.5)	2- CVCK runs another cycle 5- CVCK runs another cycle

Segment: AV5p22

After initially agreeing on an answer to the first in a series of questions (i.e. "is there flow past point bee?"), a Topic breakdown is revealed when it becomes evident that R has gone on to answer all of the questions in the series, while M believes the discussion to be focused on the very next question, apparently expecting that each question will be collaboratively discussed in turn. Once again, the only evidence of R's multiple topic transitions was his marking of answers to each question in the laboratory manual — a nonverbal display that M was clearly unaware of.

Finally, the following segments present two examples of Topic breakdowns specifically related to the lack of awareness of a partner's direction of gaze:

	M: Do I have to keep doing that? ((8)) R: eeeeew[9]www ((10))	8- After (.5) M clicks run again and both watch (1.0) 9- M makes a shrugging gesture with open hands (as in "so what?") and then gazes LB as R still stares fixedly at WS. 10- M gazes LB while R still examines WS (1.8)	8- CVCK runs another cycle, flashing arrows to show blood flow.
•	M: *uhm--looks um*mm clockw[1]ise to me (1.5) R: *Do[2] I have to write answers to these t[3]hings ohh I think we do-answer the following questions* (.9) R: *what-is the direction*-of the <u>blood</u> flow	1- R gazes LB 2- M gazes LB 3- R grabs her pen and readies it over LB	

Segment: AO2p6

• •	((2)) R: *(nn) left side* (2.6) M: *are you reading francy?* R: *h-hyeh-h-hes* M: *unkay* R: *u-<u>.hu</u>hhhh* ((3)) M: *so we haveta just* (1.0) *(dontcha) try to do[4] this ---- figure one::?* (.4) M: *unkay:: - I'm reading:: - number one - just=* [R: *hhhu+ okay* M: = *a moment[5], okay?* [R: *o*-h-ho::*kay-h-h-.hhh*(	2-Both are staring LB and reading; R is on pg2, M is on pg1 (12.6) 3- R stares LB as M audibly turns to page 2. (so both now on same page) and both read, then (11.1) R gazes WS then (2.2) drops to LB top speak. 4- M points at LB and traces back and forth across figure one as she speaks 5- M brings her finger to LB as indexer	2- Blank workspace

Segment: AO4p1

In segment AO2p6, the participants have just been discussing how to run the simulator. The Topic breakdown occurs when M, apparently believing the topic to be closed, turns to the laboratory manual to read the next question, while R continues to ponder the simulator. The divergence of topical orientation becomes apparent when M proposes an answer to the next question[11], and R's subsequent utterance clearly reveals that she had not realized that a shift in topic had occurred. The salient feature of this exchange is that the only evidence that M considers the current topic closed and is moving to the next topic is her shift in gaze to the laboratory manual; this nonverbal display is clearly not perceived by R, who realizes that M has moved to a new topic only when M eventually provides verbal evidence of her topical orientation.

In segment AO4p1, two Topic breakdowns occur in quick succession, as participants are engaged in reading the introductory paragraphs at the beginning of the laboratory manual. The first Topic breakdown is revealed when M explicitly initiates a repair by asking "are you reading francy?"; the second breakdown occurs moments later when R begins discussing the first task in the laboratory manual, prompting M to point out that she is still reading. In both cases, the Topic breakdowns occur in situations in which participants' topical orientation was evidenced only by nonverbal displays like direction of gaze, indexing the laboratory manual with a finger, and page-turning.

In sum, analysis of the circumstances under which Topic breakdowns occurred in technologically-mediated interactions revealed that breakdowns regularly occurred in the absence of strong verbal evidence of topical orientation. That is, Topic breakdowns occurred when participants relied on nonverbal resources like direction of gaze, marking of answers, and page turning to tacitly indicate that a current topic had been closed, and that discussion had moved on to a next topic. Participants in technologically-mediated interactions were clearly insensitive to such nonverbal displays, frequently resulting in the failure of their topic management efforts.

6.4.3 Copresent Topic Management

To investigate the possibility that nonverbal displays are inherently unreliable resources for topic management, the topic management behaviors of copresent participants were examined. While some Topic breakdowns did occur in the copresent condition, no consistent pattern was found between those breakdowns and the communicative resources available to participants. In particular, copresent participants displayed an intimate sensitivity to the nonverbal displays of topical orientation produced by their partners, clearly orienting to these displays to maintain shared topical orientations. For instance, consider the following exchanges:

•	M: *theres times when there is no fl*ow[3] at point B ((4)) R: well yea::h [5]---- [6]see? (.4) R: well::[7]:uhh ---- yeah right here -- *so* [M: *mhm*[8]*m* ((9)) M: *blood always flows towards the heart at point bee*[10] *or n*ot at all:::	3- R raises to WS 4- Both stare WS, then M drops to LB just before R speaks (2.2) 5- R points to G3 6- M glances WS (.4) then drops back to LB 7- R traces finger to V1, then down the connecting line to G2, pointing at gap if flows with "right here" 8- M is already marking answer in LB 9- M uses pen to point to next question, then reads (.6) 10- R raises to WS	

Segment: FF4p18

The exchange presented in segment FF4p18 demonstrates how copresent participants' apparent sensitivity to a partner's direction of gaze and manipulation of the laboratory manual was used to inform topic management. After M introduces a new topic by reading a question from the laboratory manual, R proposes an answer with the utterance “yeah -- see?” and subsequently begins to rationalize his answer by pointing to the graphs produced by the CVCK simulator. At the same time, however, M tacitly indicates that he considers the answer to be adequate and the topic to be closed by dropping his gaze to the laboratory manual and marking an answer to the question. Almost immediately after M produces these nonverbal displays, R *aborts* his rationalization, turning to gaze silently at the laboratory manual as M finishes marking the answer. This behavior clearly implies that R is sensitive to M’s nonverbal display of topic transition (i.e. shift of gaze, marking in laboratory manual), and uses these resources as a basis for interpreting M’s utterance of “mhmm” as a topic closing, rather than as an invitation to continue discussion on the current topic.

The discussion of topic management presented thus far focused primarily on the ability of one participant to perceive and successfully interpret the verbal and nonverbal displays of topic transition provided by a partner. An unfortunate effect of this tight rhetorical focus is that it fails to emphasize that shared topical orientation is collaboratively achieved phenomenon, with each participant not only making available verbal and nonverbal evidence of his or her current topical orientation, but also monitoring the displays of partners for evidence that they have perceived and correctly interpreted this evidence and are, in fact, working on the same topic. Accordingly, Topic breakdown can not be attributed solely to the insensitivity of one participant to the nonverbal evidence of topic transition made available by a partner, but also to that partner’s failure to recognize that his or her displays *have not been effective* and that, consequently, their partner remains focused on a previous topic. For example, consider the exchange previously presented in segment AO2p6. As discussed previously, the

Topic breakdown in this exchange is related to R's apparent insensitivity to M's nonverbal displays of topic transition (i.e. turning to read the next question in the laboratory manual), causing M to believe that discussion remained focused on the previous topic, while R had moved on to a new topic. However, it must be emphasized that R was equally insensitive to nonverbal evidence (e.g. M's continued gaze at the WS) that M had *failed to respond to* the topic transition signaled by R. Similar observations apply to the other examples of Topic breakdown presented earlier — in each case, the Topic breakdown occurs because of a *mutual failure* to orient to the nonverbal displays made available by a partner; one participant fails to perceive and correctly interpret the displays of topic transition provided by another, while the other participant fails to recognize that these displays have not been perceived and, consequently, is not able to take remedial action.

The exchange shown in the following segment presents a compelling example of how copresent participants were able to rely on nonverbal resources to collaboratively negotiate a topic transition, with both participants clearly orienting to the nonverbal displays of their partner:

	R: So you think they're open[4] (.2) M: Well, wait ((5)) R: Yeah (1.5) ((6)) R: Yeah (1.0) there open	4- R gazes LB, moves it in front of her, and reaches for pen ; M still gazes WS 5- R hesitates, then gazes WS. M brings cursor back to control panel 6- (7.0) M clicks on STEP again. Then four more times.	6- CVCK runs one step after another.

	((1)) R: Jim ((2)) R: Hm ((3)) R: *so-they're open at different times*	1- R gazes at LB, M continues to click on STEP; then R marks LB (8.8) then gazes WS while M continues clicking STEP (1.3) 2- M has now switched to clicking on RUN repeatedly. Both watch CVCK run (5.6) then R gazes LB (6.0) then gazes WS again (4.5). Then R glances M (.4) 3- Both gaze WS as M plays with control panel slider and then tries running a few more times. (11.2); then M finally gazes LB and R follows him after (.3). Both read for (2.4)	1- CVCK keeps running another step through the ten step cycle as M clicks. 2- CVCK now running complete pumping cycles 3- CVCK slider moves, and then CVCK runs in response to further clicks.

Segment: FF3p6

As the segment begins, the participants are discussing their answer to the question "when blood flows, are the valves open or closed," which appears on page 2 of the laboratory manual. After some further examination of the CVCK simulator, R signals her acceptance of this answer verbally ("yeah", "yeah they're open") as well as nonverbally, by directing her gaze towards the laboratory manual and marking the answer. Note, however, that R does *not go on to the next topic*, displaying a keen awareness that M is still attending to the previous topic. In particular, she is clearly aware that M has not directed his gaze at the laboratory manual and is still watching and manipulating the simulation. Finally, after giving M ample time to respond to her displays of topic transition, R verbally prompts him (i.e. "Jim") to move on to the next topic; only when M directs his gaze at the laboratory manual does the pair move on to the next topic.

This exchange clearly illustrates the value of nonverbal displays as resources for maintaining shared topical orientation and, in particular, shows how such resources are used not only to signal a change in topical orientation, but also to monitor the effect of such displays on a partner. In sharp contrast to the examples of Topic breakdown in technologically-mediated interactions presented earlier, M's behavior demonstrates an intimate sensitivity to R's nonverbal displays of topical orientation, using these displays to infer that her immediately preceding displays of topic transition have either gone unnoticed or are being ignored[12] by M, and postponing introduction of a new topic until R tacitly acknowledges the topic transition by turning to gaze at the next question in the laboratory manual.

In sum, an examination of topic management behaviors in the copresent condition revealed that copresent participants were able to use both verbal and nonverbal resources to maintain shared topical orientations. In contrast to participants in technologically-mediated interactions, copresent participants showed an intimate sensitivity towards each other's nonverbal displays of topical orientation, and were consistently able to utilize such displays to regulate their progress from topic to topic over the course of their interactions.

6.5 Discussion: Rationalizing Breakdown in Technologically-Mediated Interactions

The goal of the third and final phase of the Breakdown Analysis undertaken in this dissertation was to rationalize the significantly higher incidence of Cursor turntaking, Reference, and Topic breakdown observed in technologically-mediated interactions, by exposing constraints on certain kinds of communicative resources that were imposed by the technologically-mediated environments, weakening the evidentiary process by which participants maintain intersubjectivity. The discussion presented in the preceding three sections provides the empirical foundation for this endeavor by revealing consistent patterns in the communicative resources that were available to participants at the time that breakdowns occurred. By demonstrating that breakdowns in technologically-mediated interactions occurred when participants relied on certain kinds of communicative resources while, at the same time, showing that copresent participants were able to use these same resources to inform their interactions, the analysis strongly implies that these resources were somehow inaccessible in the audio-only or audio-video environments. In this section, we examine these resource constraints in more detail, exploring the relationship between the physical characteristics of the audio-only and audio-video environments and the empirical observations made in previous sections. To provide a foundation for this discussion, Figure 6.3 graphically summarizes the relationships between breakdowns and communicative resources exposed in the previous three sections.

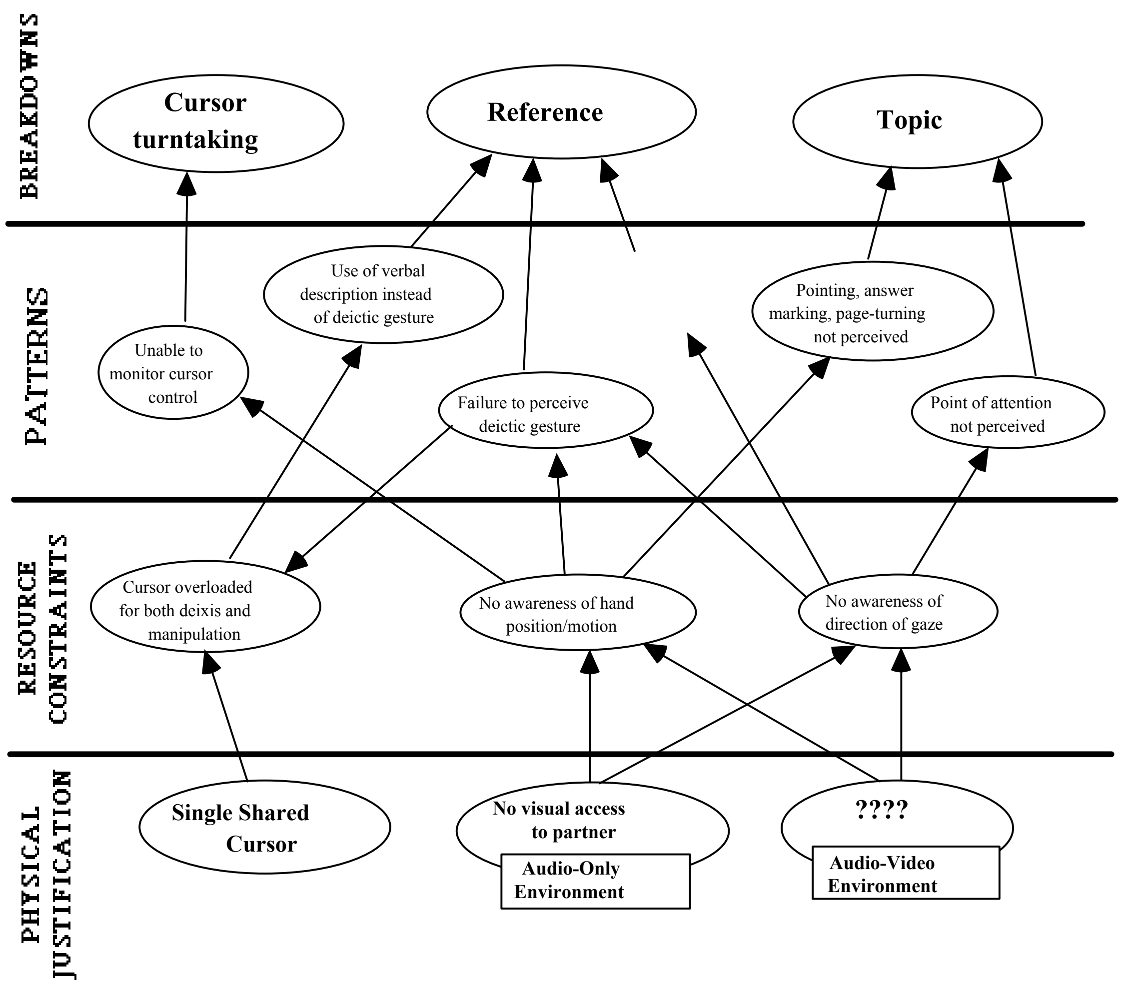


Figure 6.3: Overview of relationships between breakdowns, resource constraints, and the design of the technologically-mediated environments exposed by the qualitative analysis.

As shown in Figure 6.3, the Cursor turntaking, Reference, and Topic breakdowns that occurred in technologically-mediated interactions were related to the apparent insensitivity of distributed participants to a variety of nonverbal displays made available by their partners. Specifically, the results of the analysis for each category of breakdown can be summarized as follows:

1. Cursor turntaking breakdown. Cursor turntaking breakdowns regularly occurred in situations in which participants relied on nonverbal resources like hand

position/movement with respect to the mouse to display control over and regulate access to the shared cursor, implying that participants in technologically-mediated interaction are insensitive to these nonverbal resources for cursor management.

2. Reference breakdown. Three distinct situations were identified in which Reference breakdown occurred. First, Reference breakdowns regularly occurred when speakers failed to support references with deictic gesture, relying instead on verbal description of referents. Such breakdowns were rationalized by the observation that they invariably occurred when the shared cursor was unavailable to the speaker as a deictic tool, because it was under the control of the speaker's partner at the time. In this way, breakdowns ultimately were due to the availability of a single shared cursor. A second pattern of Reference breakdown occurred when a speaker's deictic gestures were not perceived by the listener, revealing an insensitivity of distributed participants to their partner's hand motions (either deictically or using the mouse) and current point of attention. Finally, Reference breakdowns occurred when the speaker became uncertain about the adequacy of an immediately preceding referential display, implying the speakers were unable to access nonverbal resources like direction of gaze to tacitly monitor their partner's perception and interpretation of reference.

3. Topic breakdown. Topic breakdowns were regularly associated with situations in which participants relied on nonverbal displays of topical orientation like direction of gaze, manipulation of the CVCK, and marking, pointing to, or turning the pages of the laboratory manual for topic management, rather than providing explicit verbal evidence of their topical orientations. This implies that participants were insensitive to such nonverbal displays of topical orientation.

Three common resource constraints can be seen to underlie all of the patterns of breakdown behavior summarized above: overloading of the shared cursor for both deictic and manipulatory purposes, insensitivity to a partner's hand motions or position, and the insensitivity to a partner's direction of gaze.

The insensitivity to a partner's nonverbal behaviors evidenced by participants in technologically-mediated environments strongly implies that access to these resources was somehow restricted due to the physical characteristics of the audio-only and audio-video environments. In the case of audio-only interactions, rationalizing the observed failure to utilize nonverbal resources is trivial: since no visual connection between participants was provided in the audio-only environment, the nonverbal displays of a partner were fundamentally inaccessible to participants. This observation leads to the following conclusion:

> CONCLUSION: The significantly higher incidence of Cursor turntaking, Reference, and Topic breakdown in audio-only interactions, compared to copresent interactions, can be attributed to the fact that participants had no visual contact and, therefore, were not able to access the nonverbal communicative resources made available by their partners. Lack of access to these nonverbal resources weakened the evidentiary process by which interacting participants maintained shared interpretations of action, resulting in a greater likelihood of breakdown.

The insensitivity exhibited by participants in audio-video interactions to each other's nonverbal displays is less straightforward to explain. Nonverbal behaviors like finger deixis, movement of the hand towards that mouse, and direction of gaze were all readily discernible in the remote video image available to each participant. Why, then, were participants in audio-video interactions unable to access these communicative resources to more effectively organize their interaction? The following section explores this question in more detail.

6.5.1 Remote Video Versus Physical Copresence

The qualitative analysis presented in the earlier sections of this chapter revealed that the Cursor turntaking, Reference, and Topic breakdowns observed in interactions that took place in the audio-video condition consistently occurred in situations in which participants were relying on nonverbal communicative resources like hand position, deictic gesture and direction of gaze to organize their interaction. Based on this apparent insensitivity to nonverbal behaviors, it was concluded that access to these resources was somehow constrained in the audio-video environment. At the same time, participants in the audio-video condition were provided with a large, easily-accessible video image of their partner, in which all of the nonverbal resources just mentioned were clearly available. That is, participants in audio-video interactions failed to access vital nonverbal resources, *despite* the fact that they were technically available to them in the remote video image. The issues raised by this observation can be framed in two closely-related questions:

1. Why did participants in audio-video interactions fail to access the nonverbal resources available in the remote video image to inform the evidentiary process by which they organized their interaction, reducing the likelihood of communicative breakdown?

2. Why was the remote video connection provided in audio-video condition not a functional substitute for the visual access that copresent participants enjoy?

To explore these questions, a further analysis was undertaken to investigate the way in which participants in audio-video interactions used the remote video connection, and to compare these observations to interactions that took place in the copresent condition.

As a way of characterizing the extent to which participants used the remote video image, the audio-video interactions were re-examined, counting the total number of times[13] that a participant directed gaze towards the monitor displaying the remote video image. To provide a point of comparison, copresent interactions were also examined, noting the total number of times that copresent participants directed gaze directly at a partner (i.e. turned to look directly at the person seated next to them). The results of this analysis are presented in Table 6.2.

Table 6.2: Number of gazes at partner in audio-video and copresent interactions.

Audio-Video Interactions	AV2	AV3	AV4	AV5
Number of gazes at remote video	16	86	9	23
Copresent Interactions	FF2	FF3	FF4	FF5
Number of gazes at partner	16	20	9	2

The results showed that, with one exception (i.e. AV3), participants in both audio-video and copresent interactions turned to gaze directly at their partner relatively infrequently, devoting almost all of their attention to the laboratory manual and the workspace. A statistical analysis of these results showed that there was no significant difference (U=4.0; p= 0.05) between copresent and audio-video interactions in the number of times that participants directed gaze towards the other partner.

In light of the various qualitative and quantitative results yielded by the analyses presented earlier, this observation has profound implications regarding the utility of a video image as a substitute for copresent visual access, suggesting that the access to nonverbal resources provided by a video image is fundamentally unlike copresent access to those resources. To see this, consider the following observations:

Observation 1: Nonverbal displays and breakdown. The quantitative analysis presented in Chapter V revealed that there was significantly more communicative breakdown in audio-video than in copresent interactions. Qualitative analysis revealed that this higher incidence of breakdown was related to the overwhelming insensitivity of participants in the audio-video condition to the nonverbal displays of their partners. In contrast, copresent participants were sensitive to these nonverbal resources, clearly relying on them to inform their interaction.

Observation 2: Visual access. There was no difference in the number of times that copresent and audio-video participants gazed directly at their partners. In both conditions, such gazes were relatively rare, with an overall average of 22.6 total gazes at the other participant over the course of the entire interaction; the amount of explicit attention directed at the other participant is negligible compared to the attention directed at the laboratory manual and electronic workspace.

The juxtaposition of these two observations leads to the inevitable conclusion that copresent participants were somehow able to access their partner's nonverbal displays *peripherally*, while participants in the audio-video condition were not. That is, the visual access to a partner afforded by copresence allowed copresent participants to maintain a continual awareness of a partner's nonverbal behaviors without explicitly attending to their partner. By contrast, the insensitivity to a partner's nonverbal displays evidenced in audio-video interactions strongly implies that the visual access to a partner's nonverbal displays afforded by a video image does *not* allow participants to maintain an awareness of a partner's nonverbal behavior without explicitly attending to the remote video image.

In sum, these results suggest that the access to nonverbal communicative resources provided by a remote video image is fundamentally different from that afforded by physical copresence. Whereas copresent participants were apparently able to access nonverbal behaviors like direction of gaze, pointing, and hand movement using peripheral, "back channel" (Short, Williams et al., 1976) perceptual mechanisms, participants in the audio-video condition were not. In particular, realizing access to the nonverbal resources available in a video image apparently requires a participant to explicitly attend to that video image. These observations establish the basis for the following rationale for the significantly lower communicative efficacy of audio-video interactions:

> CONCLUSION: The significantly higher incidence of communicative breakdown that occurred in audio-video interactions, compared to copresent interactions, can be attributed to the fact that participants rarely gazed at the remote video image and were apparently unable to peripherally access the nonverbal resources available in the image. Lack of access to these vital nonverbal resources weakened the evidentiary process by which participants maintain intersubjectivity, leading to a greater likelihood of breakdown.

6.5.2 Accessing Nonverbal Resources in the Remote Video Image

The analysis presented in the previous section suggests that participants in audio-video interactions were unable to peripherally access their partner's nonverbal displays in the same way that copresent participants were. In particular, the analysis implies that, though powerful nonverbal resources like hand movements, deictic gestures, and direction of gaze were available in the remote video image, participants in audio-video interactions failed to access those resources by explicitly attending to the image. This observation raises an obvious question: Why did participants in audio-video interactions not compensate for the inability to peripherally access the resources available in the remote video image by simply directing their gaze at the remote monitor more often?

An preliminary answer to this questions is suggested by the observation that, in the one interaction in which participants *did* attempt to utilize the remote video heavily (i.e. AV3, with 86 total gazes to the remote video image), the number of breakdowns in all categories was higher (see Table 5.1, Chapter V) than for the remaining three audio-video interactions. Though not a formal statistical result, this observation clearly demonstrates that merely gazing at the remote video image more frequently does not necessarily reduce the amount of communicative breakdown experienced by participants. More strongly, it implies that gazing at the remote video image to access the nonverbal resources available may actually increase[14] the overall incidence of communicative breakdown in an interaction.

To more formally explore this observation, occasions on which participants in the audio-video interactions *did* direct their gaze at the remote video image were qualitatively examined in an effort to expose communicative troubles specifically related to these attempts to utilize the remote video image. The analysis revealed that participants experienced two kinds of difficulty related to accessing the remote video image:

1. Where to look. Participants frequently displayed uncertainly over whether to look at the remote video image to perceive a partner's nonverbal behaviors, or whether to look at the workspace and laboratory manual in order to perceive nonverbal actions or interpret a partner's narrative.

2. Resolving content. The constraints imposed by the (fixed) framing and resolution limited the utility of the remote video image. Though coarse-grained phenomena like direction of gaze and hand position were readily apparent, it was impossible to read the laboratory manual, or to discern exactly what a participant was pointing at.

The following sections discuss these observations in more detail and present supporting evidence from the transcripts.

6.5.1.1 Video Schizophrenia: Deciding Where to Look

The fact that accessing the information available in a video image requires explicitly attending to that image raises a difficult dilemma for participants in audio-video interactions: the only way to access the nonverbal behaviors of a partner is by explicitly attending to the remote video image, but doing so implies not attending to ongoing action in the workspace, which is also a video image. In this way, participants were faced with a classic competition for attention situation in which they had to continually decide where to direct their attention in order to perceive the nonverbal behaviors that were "most relevant" as resources for constructing the significance of the evolving interaction. Importantly, choosing either direction of gaze rendered certain resources inaccessible — attending to the workspace or laboratory manual ruled out access to the nonverbal behaviors of a partner; attending to the remote video image made it impossible to perceive a partner's manipulations of the CVCK or to follow along in one's own laboratory manual.

As a result, participants who tried to utilize the remote video image exhibited a sort of *video schizophrenia*, snapping their gaze back and forth from one video space to the other, uncertain of which space to attend to in order to interpret ongoing talk. For example, consider the following segments:

•	R: [3]yeah it looks like *the valves* [4]*are - I would say the valves are defintely like*[5]*-o:pen* (.5) M: ehwait -- hey hey wait- [6]justin:: look [[R: (one way va) **yes ---**[7] what M: *well first of all look at the top* - we have it screwed up - these [8]aren't sposed to be here -[9]--- {aaaank} (.6) R: ooh, yeah?	3- M glances WS 4- M drops back to LB, R readies pen to mark 5- M raises to WS 6- R raises and turns to RS 7- R gazes WS, M drops to LB 8- M rolls cursor to V1u 9- M grabs V1u and drags it off towards pallete	

Segment: AV3p13

•	(.5) M: *uh*hhu -.hhh (.7) R: is this[6] a ga:::ge, right her::[7]e? ((8)) M: uhhhh that look like a+ - a:: timer or sumpin= [R: that - looks M: =uhhuhu [R: s-sort of like a gauge to me::, doesn't it? [M: ohhh okay	6- R leans and finger points to top of his WS then drops finger with "right here" 7- M turns to RS, R grabs mouse 8- M snaps to WS as R rolls cursor to gauge icon (2.0)	

Segment: AV2p13

In the exchange presented in segment AV3p13, R becomes confused about where to direct his gaze as M calls his attention to an error in their construction. As M says "hey hey wait-justin look", R snaps his gaze first to the remote video image and then, apparently realizing that M is gazing at the workspace and this is where the most relevant action is occurring, turns to look at the workspace as well. Similar confusion is evident in segment AV2p13 except that, in this case, the confusion actually causes a participant to miss a vital gesture produced by a partner. As R says "is this a gauge", he accompanies his utterance with a deictic gesture to point out the appropriate referent. However, M does not immediately realize that the deictic gesture is available in the remote video image; by the time he directs his gaze to the remote video image, the deictic gesture is no longer available since R has finished pointing. Fortunately, R uses the shared cursor to redundantly point to the referent of his utterance, avoiding potential referential confusion. The following segment shows an exchange in which participants were not so lucky:

•	(.5) M: *ye*ah:: this thing right[5] here? ----- see where the cursor is? ((6)) R: hmm? ((7)) M: *see where* the cursor is? (.7) R: yeah M: [8]isn't that sposed to be:: --- doesn't that connect those corner pieces t+	5- M rolls cursor to V in pallete, R places finger on fig1. in LB as he turns to gaze at RS 6- M snaps to gaze WS, for a moment, they are gazing each other, then R snaps to WS (1.1) 7- M snaps back to LB (.5) 8- R drops to LB	

Segment: AV2p4

In segment AV2p4, R's confusion over where to direct his gaze in order to perceive the deictic gesture associated with M's utterance "this thing right here" is strongly implicated in a Reference breakdown. As M uses the shared cursor to point out the referent of his utterance, R shifts his gaze from the laboratory manual to the remote video image, apparently expecting a deictic finger gesture. Only after M prompts with "see where the cursor is?" does M realize his mistake and snap his gaze to the workspace, allowing the referential confusion to be repaired.

The exchanges presented above illustrate the inherent ambiguity faced by participants attempting to access the remote video image, as they had to continuously decide between gazing at the remote image and gazing at the electronic workspace; the direction of gaze adopted by a participant at any one moment fundamentally determined what communicative resources were available to him or her. While attending to the remote video image provided access to certain nonverbal resources, doing so at the "wrong" moment could result in breakdown as crucial gestures or events available in the workspace were missed. In light of this inherent tradeoff, the overall communicative utility of explicitly attending to the remote video image becomes questionable.

6.5.1.2 Content: Perceiving Details in the Remote Video Image

The evidence presented in the preceding section emphasizes that there were certain "costs," measured in terms of unperceived action in the workspace, inherently associated with accessing the nonverbal resources available in the remote video image. At the same time, it was also evident that the qualitative constraints imposed by fixed resolution and framing limited the "benefits" of attending to the remote video image, by compromising the quality of the nonverbal resources it made available. For instance, consider the following exchanges:

	((9)) R: is[10] this the valve	9- R clicks RUN and they watch (2.0) 10- R leans and finger points to H1 in WS	9- CVCK runs on each click
•	((1)) M: **where** [R: this:: thing right [2]here? ((3)) M: I cant **see where you're pointing:: you retard** [R: (look where I am) ((4)) R: look where I am - the arrow (.5) M: yeah (.8) R: is that the valve?	1- R holds finger on screen, M turns to RS, then drops R drops finger and moves cursor towards H1, just as M speaks (1.9) 2- R repeats the finger point while also centering cursor on H1 3- R drops finger point as M turns back to WS (.7) 4- M snaps to WS, R is wiggling cursor over H1 throughout following sequence (.5)	

Segment: AV3p10

•	M: (nnu)- they they - they're opposite then, right? [R: yeah:: they're opposite[2] --- mine -- closes when yours op*ens*= M: *=(when) do the two valves[3] open (uuuu) close .hhh --- close at different times* ((4)) M: [5]**are you working ahea[6]::d** [R: why::	2- M clicks RUN and then drops to LB to mark 3- R marks LB, M uses pen to scan sentence as she reads quietly mumbling 4- Both mark LB then (9.7) M gazes RS, squints then speaks (2.2) 5- R raises to WS 6- R snaps to RS	2- CVCK runs
	(.7) R: **what?** (.8) M: **uhhhuh** *uhu-h-h --.hhhh- different times:: - because[1] why - .hh*hh -becu:::z (1.8)	1- R turns back to WS, adjusting hair, then drops to LB	

Segment: AV3p16

In segment AV3p10, R accompanies his question "is this the valve" with a deictic gesture, placing his finger on his workspace screen to identify the intended referent. A Reference breakdown occurs when M, who has turned to the remote video image to perceive R's gesture, is unable to discern the referent of R's utterance — a fact that she most emphatically points out to R. Another example of communicative trouble arising from the limited resolution of the remote video image is presented in segment AV3p16. In this exchange, the participants are working on answering a series of questions posed by the laboratory manual. After an extended silence, M turns to gaze at the remote video image and, upon noticing that R is still marking the laboratory manual, becomes suspicious that R has moved on to the next question; the Topic breakdown is revealed when M initiates an explicit repair by asking "are you working ahead?". In particular, the breakdown occurs when M sees R marking an answer, but it unable to discern *which* answer he is marking.

Segments AV3p10 and AV3p16 both emphasize that there is an important difference between perceiving that some nonverbal behavior is taking place, and actually being able to interpret that behavior to inform interaction. Because the quality the nonverbal resources available to participants in audio-video interactions was constrained by the resolution and framing in the remote video image, participants were often forced to use other means to somehow enhance or "repair" those nonverbal displays in order for them to be of any use. For example, in segment AV3p10, M's inability to discern the referent of R's finger deixis is resolved by providing (in parallel) a deictic gesture using the shared cursor.

6.5.1.3 Summary: Problems Using Remote Video

A question raised by the analysis presented in preceding sections is why participants in audio-video interactions did use the remote video image available to them to inform their interaction, by regularly attending to it over the course of their interaction. A qualitative analysis of what occurred when participants *did* attempt to utilize the remote video image to access the nonverbal displays of their partners presented in this section suggests several reasons why participants may have been reluctant to invest the effort required to explicitly attend to the remote video image. First, using the remote video image meant that participants had to continuously choose between looking at the shared task context, embodied in the workspace and the laboratory manual, and directing their gaze at the remote video image. In particular, attending to the remote video image at an inopportune moment could cause a participant to fail to perceive crucial gestures in the workspace, resulting in communicative breakdown. A second possible reason for the lack of interest in the remote video image is that the quality of nonverbal resources available in the image was inherently limited by the resolution and framing of the image.

Often, participants were able to perceive that nonverbal activity (e.g. pointing, answer marking) was taking place, but were not able to discern sufficient detail to allow them to interpret the significance of such behaviors.

In short, it is not clear that the communicative benefits of explicitly attending to the remote video image outweigh the costs; accessing the nonverbal resources available in the image may cause at least as much communicative trouble as it avoids. This observation explains why participants in audio-video interactions generally chose to ignore the remote video image and concentrated primarily on the workspace and laboratory manual.

6.6 Summary: Rationalizing Differences in Communicative Efficacy

In order to fully understand the differences between copresent and technologically-mediated interaction, it is important to go beyond merely *exposing* differences in the communicative efficacy to explain how those differences are related to the physical characteristics of technologically-mediated environments. Only by establishing such causal explanations can we begin to understand how existing technologically-mediated environments might be redesigned to improve the communicative efficacy of distributed interaction. Accordingly, the goal of this chapter has been to rationalize the differences in communicative efficacy between copresent and technologically-mediated interactions revealed through the stochastic comparison of breakdown presented in Chapter V. Specifically, the aim was to rationalize the significantly higher incidence of Cursor turntaking, Reference, and Topic breakdown in audio-only interactions than in copresent interactions, and the significantly higher incidence of Cursor turntaking and Reference breakdown in audio-video interactions than in copresent ones.

Because the epistemological foundation of Situated Action rules out context-independent, deterministic causal relationships between physical features of the environment and the significance of communicative behavior, a probabilistic approach based on the analysis of the communicative resources available to participants was developed. In particular, the analysis was based on the premise that, since the collaborative construction of meaning by participants is rooted in the contextual interpretation of the verbal and nonverbal displays of a conversational partner, the *likelihood* of communicative breakdown will be greater in environments in which access to these communicative resources is somehow restricted. Accordingly, analytic attention was focused on exposing consistent patterns in the communicative resources that participants were relying on to inform their interaction at the time that breakdowns occurred, implying that these resources were somehow inaccessible to participants as they worked to maintain intersubjectivity. Each category of breakdown in which significant differences between copresent and technologically-mediated interactions existed — Cursor turntaking, Reference, and Topic breakdown — was examined. The results are summarized as follows:

1. <u>Cursor turntaking</u> breakdowns were related to the insensitivity of participants to nonverbal displays of cursor control like hand position and motions with respect to the mouse and direction of gaze.

2. <u>Reference breakdowns</u> were found to be related to the availability of a single shared cursor, and to the insensitivity to a partner's direction of gaze and deictic gestures.

3. <u>Topic breakdowns</u> were related to participants' insensitivity to nonverbal displays of topical orientation like indexing or marking the laboratory manual, turning pages, direction of gaze, and manipulation of the CVCK.

Clearly, a common theme underlying the breakdowns that occurred in all three categories is an overwhelming insensitivity to the nonverbal displays of a conversational partner — communicative breakdowns regularly occurred in situations in which participants relied primarily on such nonverbal displays as resources for maintaining shared interpretations of ongoing action. At the same time, an examination of copresent interactions emphasized that copresent participants were intimately aware of the nonverbal displays of their partners, and were clearly able to use these displays to organize their interactions. These observations strongly imply that access to nonverbal communicative resources was somehow constrained in the audio-only and audio-video conditions.

In the case of audio-only interactions, the insensitivity of participants to each other's nonverbal displays is trivially explainable by the fact that participants had no visual access to their partners. More formally:

> The significantly lower communicative efficacy of audio-only interactions, as compared to copresent interactions, is rationalized by the fundamental unavailability of certain vital nonverbal displays in an environment in which conversational participants have no visual access to each other. The higher incidence of communicative breakdown in the audio-only condition was shown to be directly related to the lack of such visual access.

The reasons why participants in the audio-video condition failed to access their partner's nonverbal displays to organize their interaction were found to be more subtle. To begin with, the fact that copresent participants exhibited an intimate awareness of their partner's nonverbal displays without gazing directly at that partner implied that copresent participants were able to access nonverbal resources through peripheral, back-channel perceptual mechanisms. In contrast, the overwhelming insensitivity of participants in technologically-mediated interactions to their partners nonverbal displays implies that the remote video image did *not* support this sort of peripheral access;

participants had to explicitly attend to the video image in order to access the nonverbal resources it made available.

The inability to peripherally access the nonverbal resources in the remote video image, coupled with a failure to consistently attend explicitly to the remote video image, rationalizes the higher incidence of breakdown observed in audio-video interactions:

> The significantly lower communicative efficacy of audio-video interactions, as compared to copresent interactions, was due to a fundamental difference in access to nonverbal resources afforded by physical copresence and a remote video image. Whereas copresence affords peripheral, back-channel access to a partner's nonverbal displays, it is necessary to attend explicitly to a video image in order to access such resources. Because of the inherent tradeoffs associated with explicitly attending the remote image, the overall communicative questionable utility of doing so was questionable, explaining why participants in audio-video interactions generally chose to ignore the remote video image.

In sum, an in-depth qualitative analysis of the breakdowns that occurred in technologically-mediated interactions revealed that the significantly lower communicative efficacy in these environments was due to their failure to adequately support mutual access to participants' nonverbal communicative displays. While participants are clearly able to accomplish their communicative goals despite this constraint, the unavailability of certain vital nonverbal displays like hand motion and direction of gaze substantially weakens the evidentiary process by which participants maintain shared interpretations of action and increases that overall probability of communicative breakdown.

6.7 Notes

[1] Figure 6.1 is not meant to suggest that communicative breakdown *does not occur* in interaction in which participants have unrestricted access to communicative resources, only that it is less *likely* to occur.

[2] For the sake of brevity, pointing actions with fingers, knuckles, pencils, and laboratory manuals are not distinguished; the term "finger pointing" may be assumed to refer to all such deictic gestures, unless otherwise noted.

[3] Note that, through the use of deictic gesture to essentially *compare* two components, it becomes clear that M's utterance is meant as a critique rather than merely a mundane observation about the component in the workspace.

[4] The use of verbal description as a mechanism for negotiating shared reference in audio-only environments has been extensively explored in (Anderson, Bader et al., 1991; Clark & Schaefer, 1989; Clark & Wilkes-Gibbs, 1986; Isaacs & Clark, 1987) .

[5] Note that this breakdown is never really repaired. After M fails to respond in a timely fashion to R's indication (i.e. "help?") that she is confused about M's reference, R simply ignores M's suggestion and continues on with what she was doing.

[6] Note that M also uses verbal description to support her reference. The fact that a breakdown results anyway emphasizes the point made earlier regarding the unreliability of verbal description as a resource for establishing shared reference.

[7] Of course, participants could and often did spontaneously introduce digressionary subtopics as well, arising from the unique domain-related confusions encountered by each pair of participants.

[8] The use of questions to accomplish topic transition has been extensively documented in existing conversation analytic studies (Beach, 1993; Covelli & Murray, 1980) .

[9] For reference, the laboratory manual used by participants is reproduced in Appendix C.

[10] The question they are discussing is the last subpart to question #3 on page 3 of the laboratory manual: when does blood flow towards the heart take place at point C?

[11] The question that M is proposing an answer for is "what is the direction of blood flow" on page 2 of the laboratory manual.

[12] Note that M actually seems to be *relying* on R's ability to discern his direction of gaze to postpone the topic transition. That is, M may be fully aware that R is trying to move the interaction to the next topic but is actually utilizing his direction of gaze to make it clear to R that he is not ready to move on.

[13] A better measure of utility might have been total time spent gazing at the remote video image. Unfortunately, the landmark-based transcription notation (see Chapter 4) used for this study did not capture temporal features of the interaction in enough detail to allow this.

[14] This empirical observation supports long-standing claims by social psychologists (Short, Williams et al., 1976) that the value of video-mediated telecommunication has been consistently overrated.

CHAPTER VII

SUMMARY AND CONCLUSIONS

In the past decade, substantial decreases in the cost of network bandwidth, coupled with an overall increase in network connectivity and robustness, have led to an explosion of interest in sophisticated technologically-mediated communication environments that enable widely distributed participants to collaboratively accomplish their communicative and creative goals. As discussed in Chapter I, a variety of communications environments for desktop conferencing, group interaction, and distributed design have been developed using technologies ranging from mundane typed-text to powerful audio-video environments to sophisticated virtual realities. This technology has moved from the research laboratory into the public sector in recent years, with a growing number of network-based communications environments becoming available on the commercial software market. Judging by this trend, it seems clear that technologically-mediated communications environments will play an increasingly important role in modern society, fundamentally changing the way in which we interact both personally and in professional settings.

Unfortunately, the aggressive pace of technical development has far outstripped our understanding of how these novel communication environments affect the *quality* of the interactions they enable. Ultimately, the goal of any technologically-mediated environment is to present users with a simulacrum of copresent interactions that allows participants to accomplish their communicative goals as easily and efficiently as if they were interacting face-to-face. That is, the communicative efficacy of technologically-mediated interactions should be the same as copresent interaction. This observation frames the central research issue addressed in this dissertation:

> Research Issue: To what extent is technologically-mediated interaction functionally equivalent to copresent interaction? How does the communicative efficacy of technologically-mediated interaction compare to that of copresent interaction?

The only way to address this issue is to operationalize the notion of communicative efficacy by somehow characterizing the extent to which an environment supports successful communication. One reason that this has proven to be a difficult problem is that there is no deterministic formal model for communicative success. Where designers of a new rocket, for instance, can rely on the laws of physics, both to describe the system's behavior predictively and to rationalize it in retrospect, no such model has been developed for human communication in general. In this vacuum, the development of the current crop of computer-mediated environments has largely been driven by and oriented around the technical challenges posed by distributed interaction. By focusing on issues like bandwidth, frame rate, color depth, and sampling rate, these projects make the tacit assumption that "more is better" — that higher bandwidth and better resolution inevitably lead to a higher communicative efficacy. Clearly, this approach places form before function, ignoring functional utility of the environment in favor of abstract parameters.

A number of studies have attempted to remedy this shortcoming by empirically comparing copresent and technologically-mediated interaction based on metrics like user satisfaction (Isaacs, Morris et al., 1995; Olson, Olson et al., 1995; Tang, Isaacs et al., 1994) , quality of work (Olson, Olson et al., 1995) , and task-activity structure (Olson, Olson et al., 1995; Tatar, 1989) . Though all of these approaches provide a basis for asserting that interactions in one environment have a higher communicative efficacy than in another, they yield no insights as to *why* differences in efficacy exist. For instance, user satisfaction surveys can tell us that users prefer one communication environment over another, but do not reveal the communicative difficulties experienced by participants in a “less satisfying” environment that are presumably the root cause of their dissatisfaction. This limitation arises from the fact that metrics like user satisfaction, quality of work, and task-activity structure characterize the communicative efficacy of interactions *indirectly*, inferring the amount of communicative difficulty experienced by participants from the overall outcomes or structure of interactions.

The study presented in this dissertation focuses analysis directly on the communicative interaction of participants, assessing the communicative efficacy of interactions by documenting the number and nature of communicative confusions, or *breakdowns*, experienced by participants. The research contributions of this dissertation are summarized as follows:

Research Contribution: Methodology

The methodology of Breakdown Analysis was developed to explore the integration of the qualitative methodologies of Conversation and Interaction Analysis with more traditional quantitative techniques used in the hard sciences. Unlike the existing empirical techniques mentioned above, Breakdown Analysis works to directly assess the quality of participants’ interaction by documenting the number of communicative breakdowns they experienced. Specifically, the analysis was based on a comparison of the number of Verbal turntaking, Cursor turntaking, Topic, and Reference breakdowns experienced by participants in the three environments.

An important advantage of this approach is that it yields an explicit and concise characterization of the communicative troubles encountered by participants in environments with relatively poor communicative efficacy, providing a strong foundation for a focused investigation of *why* more breakdowns occurred in these environments. By articulating causal relationships between the physical characteristics of an environment and the communicative troubles experienced by users, the analysis establishes a solid basis for future redesign.

<u>Research Contribution: Differences in Communicative Efficacy</u>

The work presented in this dissertation addresses the primary research issue stated earlier by directly comparing the communicative efficacy of copresent and technologically-mediated interaction. Specifically, the study compared the communicative efficacy of the copresent condition to that of two technologically-mediated environments that are representative of the technologies used in many existing systems: an audio-only environment and an audio-video environment. The analysis yielded two major results:

<u>Result:</u> *The communicative efficacy of technologically-mediated environments was substantially lower than that of the copresent condition.* Participants in audio-only interactions experienced significantly more Cursor turntaking, Reference, and Topic breakdown; audio-video interactions showed significantly more Cursor turntaking and Reference breakdown. In no category was there significantly less breakdown in technologically-mediated interactions than in copresent ones.

<u>Result:</u> *There was no difference in communicative efficacy between the two technologically-mediated environments.* No significant differences in the amount of breakdown was found for any of the four categories between the audio-only and the audio-video condition. This result is somewhat surprising and runs contrary to the intuition that an environment that provides a higher bandwidth connection between participants necessarily enhances their ability to communicate effectively.

Research Contribution: Rationalizing Breakdowns

The concise characterization of communicative trouble yielded by Breakdown Analysis made it possible to explain why observed differences in communicative efficacy exist. By establishing causal relationships between breakdowns and the physical characteristics of the environment, the analysis shows how the technologies used to implement the two distributed environments impinged on their communicative efficacy. Specifically, the analysis yielded the following results:

Result: *Access to nonverbal displays is vitally important for organizing interaction and avoiding breakdown.* The analysis shows that breakdowns that occurred in technologically-mediated interactions were related to a gross insensitivity to nonverbal displays like direction of gaze, deictic gesture, and manipulation of objects in the workspace. Breakdowns occurred in situations in which participants were relying primarily on such displays to make available their current interpretation of ongoing action. On the other hand, copresent participants were observed to be extremely sensitive to each other's nonverbal displays, using them to inform their cursor, topic, and reference management activities. This result clearly demonstrates the value of visual contact as a resource for organizing interaction.

Result: *A video image is not a functional substitute for copresent visual access.* Despite the fact that participants in audio-video interactions were provided with a high quality video image of their partner, they were unable to utilize this image to access the nonverbal displays of their partners. A detailed analysis of how participants in audio-video interactions used the remote video image revealed profound pragmatic differences in the access to a partner's nonverbal displays afforded by the remote video image and that afforded by physical copresence. While copresent participants were able to monitor each other's nonverbal displays using peripheral perceptual mechanisms (i.e. without gazing directly at their partner), the remote video image did not afford the same kind of access. Perceiving the nonverbal displays available in the remote video image required participants to focus their attention explicitly on the image. As a result, participants in audio-video interactions had to continually split their attention between the workspace and the remote video image, frequently leading to additional breakdown. Moreover, the resolution and framing of the remote video image made it impossible for participants to discern fine-grained details in the remote video image, e.g. exactly what a partner was pointing at, substantially limiting its utility.

In sum, the analysis presented in this dissertation shows that the two technologically-mediated environments *did not* provide simulacrums of copresent interaction that were functionally equivalent to physical copresence; communicative efficacy in the technologically-mediated interactions was consistently lower than in the copresent condition. From the standpoint of design, the most important result yielded by this study is that the availability of a video channel in the audio-video condition did not contribute to the communicative efficacy of that environment; the audio-video condition was functionally equivalent to the audio-only condition. This observation clearly has profound implications for technologically-mediated environments currently being made available to the public, most of which provide participants with a video image. Given that participants engaged in task-oriented interactions are unable to take advantage of this resource to better organize their interaction, it might be sensible to find better ways to utilize the bandwidth devoted to the remote video image. The following section speculates on several ways in which this might be done.

7.1 Implications for Future Design

The results yielded by this research suggest that the role of video in the future development of technologically-mediated environments must be carefully reconsidered. On the one hand, the conclusion that nonverbal displays are crucial for avoiding breakdown implies that any technologically-mediated environment that hopes to match the communicative efficacy of copresent interaction must somehow make such displays available to participants. On the other hand, the observation that audio-video participants were unable to effectively utilize the remote video image demonstrates that a video image is not a functional substitute for copresent visual access. These observations are summarized in the following design prescription:

> Design prescription: In order to match the communicative efficacy of the copresent condition, a technologically-mediated environment must somehow make the nonverbal displays of participants mutually available.
>
> Corollary: A video image displayed on a traditional monitor does not provide effective access to such nonverbal displays.

In general, there are two solutions to the quandary posed by this prescription: explore alternative modes of visual access that overcome the limitations associated with monitor-based video, or provide alternative communicative resources to compensate for the constrained access to nonverbal displays.

7.1.1 Exploring More Naturalistic Visual Access

The qualitative analysis of the breakdowns that occurred in audio-video interactions clearly demonstrated that the remote video image was not functionally equivalent to copresent visual access; participants were unable to use the remote image to access each other's nonverbal displays. One solution to this problem is to develop alternative visual representations that make remote visual access more similar to copresent visual access. To understand what this might entail, it is important to succinctly articulate the functional differences between a video image and copresent visual access. The analysis presented in the latter sections of Chapter VI suggests that there are at least three important differences:

1. Point of view. Because copresent participants are both *embedded* within the same 3-dimensional physical space, they have equal and complete visual access to that space. By contrast, the camera and video monitor used to implement a remote video image establish a *detached* 2-dimensional point of view, providing a fixed-frame "porthole" into a partner's visual space. The difference between the embedded and detached point of view can be likened to the difference between looking into a room through a window and actually being in the room. A person inside the room has full access to the three dimensional space within which he or she is embedded, and retains a peripheral auditory and visual awareness of the objects (or conversational participants) within the room, even when not gazing directly at them. The perspective of a person looking into a room through a window, on the other hand, is fundamentally constrained by the visual space framed by the window. The same is true of a remote video image (see Figure 7.1): because of the fixed framing of the camera, the remote participant has access only to the visual space displayed in the video image and therefore has no real sense for the overall layout of the 3-dimensional space inhabited by a partner.

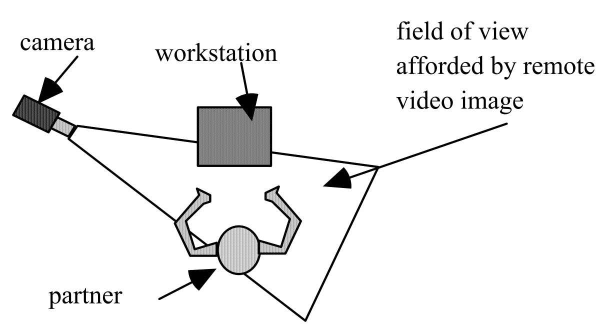


Figure 7.1: Top view of a partner's work area showing the tightly constrained field of view available to a second participant via the remote video image.

A particularly important consequence of the detached point of view is that it essentially creates two distinct Cartesian spaces — one for the observer and one for the subject. This makes interpretation of deictic gestures and spatial references, e.g. “in front of,” “to the right of,” produced by a remote partner more difficult, since the observer must essentially map between the two Cartesian spaces in order to construct the significance of these productions. This may be one reason why participants in the audio-video condition had difficulty determining what their partners were pointing at, even when they were gazing directly at the remote video image.

2. Resolution and scaling. Another important limitation of a video image is that the constraints imposed by the video format, i.e. the number of pixels supported, inherently constrain the amount of detail available in the image. As the field of view captured in the video image increases, the resolution of the image decreases. To make matters worse, the remote video image usually has to be reduced in scale in order to fit a reasonably large field of view onto the remote video monitor. The end result is that an observer’s ability to discern fine-grained details in the remote video image is severely compromised. The analysis presented in Chapter VI revealed that this limitation was a significant impediment to participants in the audio-video condition. For example, participants were able to see that a deictic gesture was taking place, but were not able to discern what object was being pointed at.

3. Perceptual mechanisms. As a result of their embedded point of view, copresent participants were able (see Figure 7.2) to rely on low-resolution *peripheral* perception to maintain a continual awareness of a partner’s nonverbal displays, while keeping their high-resolution *focal* vision trained on the shared workspace. Because of the poor resolution and scaling of the remote video image, however, participants in audio-video interactions were not able to access the nonverbal displays available in the remote video

image using these same peripheral perceptual mechanisms. Rather, participants had to explicitly train their high-resolution focal vision on the remote video image in order to perceive nonverbal displays. As discussed in Chapter VI, this resulted in a situation in which participants had to continually split their attention between the workspace and the remote video image, frequently leading to breakdown when participants gazed in the "wrong" direction and missed vital nonverbal displays.

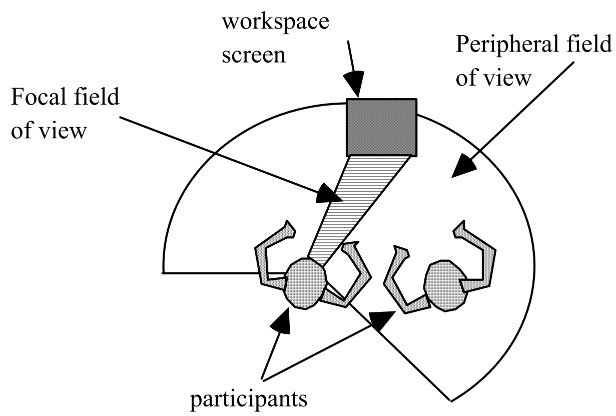


Figure 7.2: Top view of two copresent participants showing the narrow focal field of view (focused on the workspace) and the broad peripheral field of view used to maintain awareness of a partner's nonverbal displays.

In sum, the point of view, resolution, and scaling of a video image make it fundamentally different from copresent visual access. Because of these differences, audio-video participants were not able to rely on peripheral perceptual mechanisms to track each other's nonverbal displays. These observations suggest that any visual representation that is functionally equivalent to copresent visual access must meet three criteria:

1. The representation must support an embedded point of view, somehow placing both participants within the same visual space.

2. The representation must present the remote space at a natural (1:1) scale and at a resolution that preserves fine-grained details.

3. The representation must support peripheral visual access[1] to the behaviors of a collaborating partner.

In recent years, several systems have been developed that meet these criteria to some extent. For example, Clearboard presents distributed participants with the illusion that they are seated face-to-face, separated by a pane of glass on which both participants are able to draw using felt markers (Ishii & Kobayashi, 1993) . The MAJIC system is slightly more elaborate, supporting the interaction of up to four participants; each participant is presented with the illusion that conversational partners are seated in a semi-circle opposite them (Okada, Maeda et al., 1994) . By carefully arranging the video cameras that record the remote images and controlling where participants sit, both MAJIC and Clearboard support a limited form of mutual eye gaze.

Though both of these systems appear to satisfy the three criteria laid out above, they suffer from a common limitation: participants always appear opposite of each other. Though this constraint presents no problem in mundane conversations, it is bound to lead to problems in task-oriented interactions in which participants are referring to objects, e.g. a sketch, in the space "between" them. Since there is no way for participants to more closely synchronize their points of view by sitting side-by-side, the significance of spatial references like "to the right of" and "in front of" will differ[2] from participant to participant.

Systems like Clearboard and MAJIC both work to support an embedded point of view by physically expanding the shared visual space to encompass both participants. Recently developed *virtual reality* technologies take this idea to an extreme by essentially removing participants from their physical surroundings and bringing them together in a shared virtual space. For example, the DIVE system allows any number of participants to meet in a virtual conference room, providing a virtual whiteboard to support task-oriented discussions (Benford, Bowers, Fahlen, Greenhalgh, & Snowdon, 1995) . Because there are no constraints imposed by camera positioning or framing, participants in the virtual space are able to move about to establish whatever point of view is convenient. However, this flexibility is not without cost: the extremely high computational requirements of virtual reality systems severely limit the amount of detail that can be represented in the virtual world. Every time any participant in the virtual space makes a move, the virtual scenes presented to each participant must be recomputed to reflect the change. The only way to manage this computational complexity is by radically simplifying the representation of the virtual space: participants are mapped onto low-resolution three-dimensional models that reflect a participant's overall motions and body orientation but do not present a photo-realistic image of the participant. The fidelity of the virtual space is also constrained by the quality and quantity of the sensors attached to participants. For example, the DIVE system tracks only the position of the head and hands; features like facial expression, mouth movements (e.g. while talking), and so on are not represented. Because of these fidelity constraints, the access to nonverbal displays afforded by a virtual reality may be just as inadequate as that afforded by a monitor-based video image. That is, though virtual reality systems overcome the point of view restrictions inherent in a video image, these gains may be nullified by the poor fidelity of the representation.

In sum, one approach to addressing the insensitivity to nonverbal displays observed in this study is to explore alternative visual representations that provide more realistic simulacra of copresent visual access. However, though systems like Clearboard, MAJIC and DIVE appear to overcome the most serious limitations of monitor-based video images, each of these approaches introduces new and unique limitations of its own.

7.1.2 Providing Compensatory Resources

While it remains to be seen whether sophisticated systems like Clearboard or MAJIC succeed in supporting more natural, embedded access to a partner's nonverbal displays, it is clear that all such systems suffer from a major drawback: they are relatively elaborate, expensive, and difficult to set up and maintain. In other words, they do not represent a *practical* solution that can be readily incorporated into the design of today's desktop workstations.

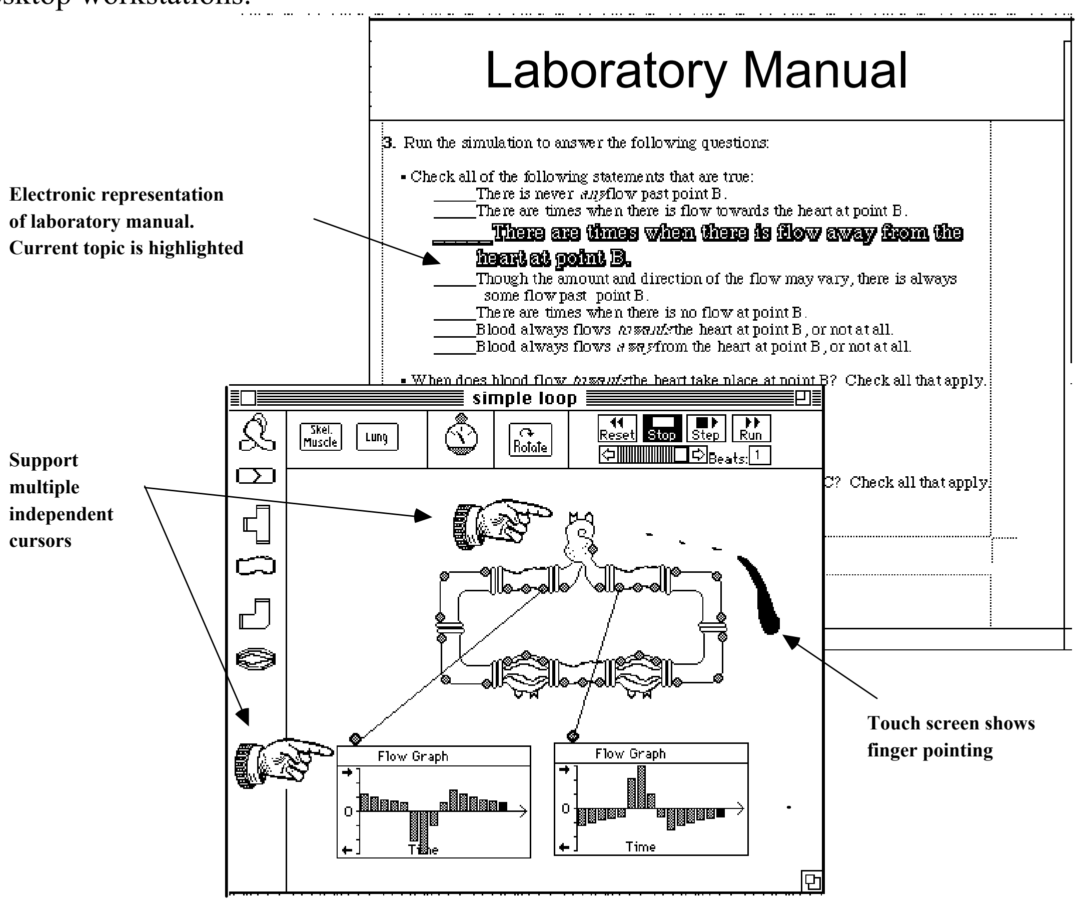


Figure 7.3: The CVCK interface with compensatory resources.

Rather than working to improve the simulacrum of visual copresence perceived by participants, a more pragmatic approach is to think about how one might augment or, indeed, replace the remote video image by providing additional *artificial* resources that compensate for the inaccessibility of a partner's nonverbal displays. The underlying premise of this approach is that it is not the video image itself that is important, but rather the nonverbal displays that visual access makes available. If we can find alternative ways of representing these nonverbal displays, then it should be possible to reduce the incidence of breakdown.

To illustrate this idea, Figure 7.3 shows how the audio-video environment investigated in this study might be modified to provide such compensatory resources.

As shown in Figure 7.3, one way to compensate for participants' inability to perceive nonverbal topic management displays like pointing to the laboratory manual, marking answers, and turning pages might be to electronically present the laboratory manual in the shared workspace. Participants could make available their topical orientation by using the cursor to point at the electronic laboratory manual, highlighting the current topic of discussion, marking answers, scrolling to see the next question, and so on.

One might compensate for distributed participants' insensitivity to a partner's finger pointing by using a touch screen to make such deictic gestures available in the shared workspace. When a participant points to their workspace screen, the system could respond by producing a "fingerprint" in the shared workspace to mark the location of the pointing party's finger. The mark could be made to fade rapidly after the pointing finger is removed; moving the finger across the workspace would produce a fading smear that naturally indicated the speed and direction of the gesture. For example, the dark smudge shown in Figure 7.3 would result if a participants used their finger to trace a clockwise direction of blood flow around the cardiovascular loop.

Finally, an obvious way to compensate for participants' inability to access nonverbal displays like hand position and direction of gaze to regulate access to the shared cursor is to simply provide each participant with an independent cursor, allowing both participants to simultaneously gesture or manipulate the workspace. Though current operating systems do not provide general-purpose support for this capability[3], several existing technologically-mediated environments provide independent cursors for each participant (Bly & Minneman, 1990; Minneman & Bly, 1991; Tang & Minneman, 1990) .

In general, the goal of these modifications is to eliminate the need for remote visual access to a conversational partner, either by avoiding certain organizational requirements altogether, e.g. by providing a second cursor, or by explicitly representing nonverbal displays that were formerly available only in the remote video image within the shared workspace. As a result, participants would have access to these communicative resources without having to explicitly divert their attention to a remote video image.

A final advantage of concentrating all nonverbal displays in the shared workspace is that it reduces the importance of tracking a partner's direction of gaze. As discussed in Chapter VI, one of the main advantages to having an awareness of a partner's direction of gaze was that it allowed copresent participants to notice when their partners *were not attending* to various nonverbal displays. For example, speakers were able to notice that their partners were gazing at the laboratory manual at the time that the speaker produced a deictic gesture and, as a result, deduce that the gesture had not been perceived. The same was true for gestures like pointing and gesturing with respect to the laboratory manual — participants were able to monitor their partner's direction of eye gaze to determine whether the gestures had been perceived. Based of this awareness, copresent participants were able to take remedial action, explicitly drawing a partner's attention or producing redundant displays before continuing on with the conversation. By contrast, breakdowns in technologically-mediated interactions frequently occurred because participants were insensitive to their partner's current point of attention and were therefore unable to regulate their interaction in this way. For example, Reference breakdowns occurred when participants failed to notice that a partner was gazing at the laboratory manual and, consequently, failed to perceive a deictic gesture using the shared cursor.

By modifying the system as discussed above to represent the laboratory manual and deictic finger gesture in the shared workspace, participants' attention is focused exclusively on the workspace. That is, there is no longer any reason for participants to be gazing at anything *but* the shared workspace; participants can confidently assume that a partner is gazing at the shared workspace at any given point during the interaction and will perceive all nonverbal displays.

In sum, one way to overcome the inability of participants to access the nonverbal resources available in a remote video image might be to provide alternative, more easily accessible mechanisms for regulating interaction. Recent work by Dykstra-Erickson et al. (1995) shows that participants are able to adapt their communicative practices as they gain experience in a technologically-mediated environment. This suggests that participants could learn to utilize artificial organizational resources like the ones described above to compensate for the inaccessibility of nonverbal displays contained in the remote video image. Because these approaches work to represent nonverbal cues available in the remote video image in the shared workspace, they could potentially eliminate the need for the relatively costly (in terms of bandwidth) video channel.

7.2 Future Studies

Ultimately, the only way to determine whether the modifications discussed in the preceding section actually succeed in supporting more effective access to a partner's nonverbal displays is by implementing them and empirically comparing the resulting environments to copresent interaction using Breakdown Analysis. For example, the study presented in this dissertation could be repeated, replacing the remote video monitor with a system similar to Clearboard. Other design variations like reduced video frame rate, lower quality audio, and so on might be explored as well. More generally, the methodology of Breakdown Analysis is ideally suited for assessing the communicative performance of any technologically-mediated environment, making it a powerful tool for expanding our understanding of which technologies are most appropriate for specific kinds of communicative endeavors (Teasley, Olson, & Meader, in preparation) .

The following two sections discuss two communicative scenarios that differ substantially from the one investigated in this study and warrant special attention in future work: Scenarios in which participants are engaged in personal interactions, and scenarios in which participants have substantial previous experience interacting in the technologically-mediated environment. Finally, Section 7.2.3 explores the prospects for streamlining the methodology of Breakdown Analysis to make it more practical for evaluating designs in the modern fast-paced world of industrial software development.

7.2.1 Breakdown in Personal Interactions

The interactions investigated in this study were distinctly task-oriented in nature. Participants were given a specific, well-defined task to accomplish and devoted their attention exclusively to collaboratively manipulating the CVCK simulator to accomplish that task. This tight focus on producing a tangible product makes task-oriented interactions fundamentally different from more socially oriented *personal* interactions like contacting a friend to see if he or she would like to have lunch, discussing a movie, or debating the merits of a memo sent around by a coworker. Specifically, there are at least three important differences between personal interactions and task-oriented interactions:

1. No predefined topic structure. The overall topic structure defined by the task solution process (e.g. the laboratory manual in the CVCK task) does not exist in personal interactions. Rather, new topics of conversation are dynamically generated by participants and topic transitions explicitly negotiated as the conversation evolves.

2. Lower emphasis on physical object reference. In task-oriented interactions, discussion is tightly focused on some mutually available workspace containing the task representation. As a result, participants must establish and maintain shared reference to the physical objects and entities within that workspace. For example, to discuss the manipulation and behavior of the CVCK simulator, participants continuously referred to various cardiovascular components and their spatial location within the workspace. In comparison, the amount of physical object reference that occurs in personal interactions is relatively small. For example, two participants engaged in a mundane conversation about what they did over summer vacation will almost certainly produce fewer references than if they were discussing the CVCK simulator.

3. No need to manipulate or gesture. Since participants in personal interactions are not working to collaboratively accomplish a specific task, they are not required to regulate access to a shared cursor or gesture at a mutually available representation. Indeed, there is no need for a shared workspace at all.

Collectively, these differences imply that a Breakdown Analysis comparing copresent to technologically-mediated interaction would yield substantially different results for personal interactions than for task-oriented ones. For instance, the analysis presented in Chapter VI showed that Topic breakdowns generally occurred in the absence of explicit verbal topic management, when a participant implicitly moved on to the next topic defined by the laboratory manual. Because new topics are always verbally introduced in personal interactions, it is reasonable to assume that the likelihood of Topic breakdown will be reduced. Similarly, the fact that participants do not have to continuously establish direct reference to objects in some mutually available task context suggests that Reference breakdown will be much less of a problem in personal interactions. Finally, Cursor turntaking breakdown would not be an issue at all in personal interactions, since there is no need for a shared cursor.

In short, it is not clear that the categories of communicative breakdown that were identified in the analysis presented in this dissertation even exist as consistent patterns of communicative trouble in personal interactions; entirely different categories of breakdown may need to be developed. Consequently, the conclusions regarding the relative communicative efficacy of copresent and technologically-mediated interaction yielded by this study can not be assumed to apply to personal interactions. This motivates a future Breakdown Analysis to explore the differences between copresent and technologically-mediated interactions in which participants are engaged in mundane personal conversation.

7.2.2 Breakdown with Experienced Participants

As pointed out in Chapter V, there was no decreasing trend in the number of breakdowns experience by participants over the course of interactions; the incidence of breakdown was not consistently lower in the closing phases of interactions than in the opening phases. At the same time, it has been observed that participants who regularly use a technologically-mediated environment are able to adapt their communicative practices as they gain experience in the environment, developing novel communicative mechanisms to take advantage of the unique characteristics of the environment (Dykstra-Erickson, Rudman et al., 1995) . This suggests that distributed participants may, in fact, *eventually* develop compensatory mechanisms that reduce the incidence of breakdown in their interactions; the interactions examined in this study were simply not long enough for such adaptation to take place.

The only way to conclusively resolve this issue is to design a longitudinal study that compares the amount of communicative breakdown experienced by participants at various points in time, as they gain experience with a particular technologically-mediated environment.

7.2.3 Can Breakdown Analysis be Streamlined?

A central claim made in this dissertation is that the methodology of Breakdown Analysis represents a powerful analytic tool for exploring the differences in communicative efficacy that exist between copresent interaction and interaction in technologically-mediated environments, including both existing systems and those to be developed in the future. For instance, the two future studies suggested in the preceding sections both rely on Breakdown Analysis to expand our understanding of how the communicative efficacy of technologically-mediated environments varies with respect to the experience level of participants and the type of task they are engaged in. One question raised by the prospect of regularly using Breakdown Analysis to assess the communicative efficacy of new environments is whether the methodology can somehow be streamlined to transform it into a more practical analytic tool. This section briefly discusses the costs of performing a Breakdown Analysis, and speculates on how these costs might be reduced in future studies.

A convenient way of characterizing the cost of any evaluative technique based on exploratory sequential data analysis is by determining its ratio of analysis time to session time (AT:ST) (Sanderson & Fisher, 1994) . For example, if a given methodology requires ten hours of analysis to process a single hour of data (i.e. an hour of videotaped interaction), then it has an AT:ST ration of 10:1. As a starting point for this discussion, Table 7.1 uses this schema to summarize the effort required for Breakdown Analysis.

Table 7.1: An overview of effort required for Breakdown Analysis expressed in terms of AT:ST ratios.

	Estimated AT:ST ratios for Breakdown Analysis
Study #1: Identifying patterns of breakdown	60:1
Study #2: Quantitative Analysis	2:1
Study #3: Rationalizing differences in efficacy	30:1
TOTALS	**92:1**

As indicated in Table 7.1, performing a Breakdown Analysis requires a substantial investment in effort: the total amount of time required to perform an analysis like the one presented in this dissertation[4], which consisted of 12 approximately half hour sessions, was roughly 552 hours, or about three and a half months of intensive labor. The two qualitative studies are by far the most effort-intensive components of the analysis, requiring extensive and painstaking analysis of videotape or transcript data. By comparison, the effort required for the quantitative analysis is quite modest.

More generally, the cost of performing a Breakdown Analysis is relatively high in comparison to similar techniques based on the qualitative analysis of videotape data, which typically have AT:ST ratios between 5:1 and 50:1 (Sanderson & Fisher, 1994) . This stems from the fact that Breakdown Analysis in based on the in-depth and often iterative analysis of videotaped and transcribed data, and that such analysis is inherently effort-intensive.

In light of this observation, it is clear that the only way to significantly reduce the effort required for a Breakdown Analysis is by exploring ways of reducing the raw amount of videotape data to be analyzed. There are two[5] ways in which this might be done:

1. Reduce the time required to complete the given task. The CVCK task required, on average, approximately a half hour to complete. Perhaps one could find a task that requires less than ten minutes to perform.

2. Reduce the number of environments compared. Focus the analysis on interactions in only two — or possibly just one — communication environment.

Though each of these modifications would produce the desired reduction in effort required to perform the Breakdown Analysis, they may also compromise its integrity. For example, as a derivative of Interaction Analysis, Breakdown Analysis is centered around the examination of *naturally-occurring* interactions. That is, the interaction that participants engage in must closely match a real world interaction that participants might normally engage in outside of the laboratory. Finding a naturally-occurring, non-trivial collaborative activity that requires less than ten minutes to perform may be difficult. Another problem with reducing the session time is that the average rate at which breakdowns in some categories occur is quite low; very short sessions may not allow enough time for a meaningful number of breakdowns to occur, short-circuiting the analysis. For example, the total number of Topic breakdowns per session (see Table 5.1, Chapter V) ranged between one and four per session. If sessions were shortened to less than ten minutes, there would undoubtedly be many sessions in which no Topic breakdowns occurred. Indeed, in such short sessions, it might be impossible to ask participants to cover more than one or two topics.

Perhaps the most promising way to reduce the effort required for a Breakdown Analysis is by reducing the number of environments that are compared. At the very least, the number of environments could be reduced from three to two; the reasons for comparing three environments in this study were entirely pragmatic (i.e. motivated by the desired to compare representative examples of existing technologies) rather than methodological. By comparing only two environments, the cost of the analysis would be reduced by a third.

An even more intriguing possibility is to reduce the number of environments even further, analyzing interactions in only a single technologically-mediated environment. Since the central issue addressed by a Breakdown Analysis is how the communicative efficacy of a novel technologically-mediated environment compares to the copresent condition, it might be possible to establish "breakdown benchmarks" for copresent participants performing a variety of well-defined tasks, and then use these benchmarks as a point of comparison when evaluating new environments. That is, technologically-mediated environments could be evaluated by choosing the benchmark task that most closely matches the intended use of the environment, documenting the breakdowns experienced by participants as they collaboratively perform that task in the environment, and comparing the results to the copresent benchmark for that task. In this way, the analysis of copresent interactions for each Breakdown Analysis could be eliminated, reducing the cost of the analysis by another third. Even with a relatively high AT:ST ratio of 92:1, this drastic reduction in the number of hours of videotape data examined by the analysis would result in an overall analytic effort comparable to existing empirical techniques (Sanderson & Fisher, 1994) .

Though tempting from a practical standpoint, the benchmark approach raises a number of important theoretical concerns. The underlying assumption of any evaluative approach based on benchmarks is that the environmental conditions, or context, in which each new design is tested can be made to be identical to the context that existed when the benchmarks were established. In most domains this is non-problematic. In evaluating the performance of a new compiler, for example, the analyst must simply ensure that the compiler is installed on the same type of machine and in the same operating environment used to establish the benchmark. Unfortunately, the inherently situated character of human communication makes this condition of "same context" fundamentally unattainable for communicative interactions, since the contextual features relevant to the interpretation of a given communicative display can never be completely enumerated and are unique to each new situation. More concretely, contextual factors like the background and experience of participants and participants' level of expertise in the task domain are impossible to succinctly define and, therefore, impossible to control for. At the same time, it is reasonable to presume that, by working to *minimize* these sources of variability, it might be possible to create a useful system of benchmarks; the following constraints establish a framework for designing benchmark tasks:

1. The benchmark tasks should require a minimum of domain-specific knowledge to complete. The goal of this constraint is to equalize the amount of expertise that current and future participants bring to bear on the task. In this respect, the CVCK task used in this work represents a counterexample to the sort of task one would want for a benchmark, since it would be difficult to ensure that participants in future subject pools

had no more or less knowledge of cardiovascular function that those used to establish the benchmark. An exception to this constraint is when the class of systems being designed is highly specialized (e.g. systems to support collaborative architectural drawing). In such cases, the benchmarks should be set by domain experts performing domain-related tasks.

2. Careful attention should be paid to the details of the copresent condition; different benchmarks should be developed for even small differences. For example, the results of the analysis presented in this dissertation imply that the fact that participants were sitting side-by-side rather than opposite each other was relevant to their ability to perceive nonverbal displays. Another difference already mentioned is the one between task-oriented interactions and those aimed at mundane social interaction. Other differences might include positioning of the shared workspace with respect to participants, the means provided for manipulating the shared workspace and so on. Each of these differences should be reflected in the benchmarks, with a separate benchmark established for each set of conditions.

In sum, there are at least two ways in which Breakdown Analysis might be streamlined, making it a more practical methodology for everyday use. First, it might be possible to find a more compact task for participants to collaboratively perform. In this way, the amount of time for each session would be reduced, leading to an overall reduction in the amount of videotape data to be analyzed. Second, one might reduce the number of environments compared to two, or possibly even just one, comparing interactions in a single environment to a previously established "benchmark". Though both of these approaches show some promise, it is not entirely clear that they can be successfully implemented without compromising the integrity of the methodology. This open issue motivates several future studies focused on the methodology of Breakdown

Analysis itself. For example, one study might investigate the effects of varying the length and complexity of the task on the outcomes of the analysis. A subsequent Breakdown Analysis might then be performed comparing copresent interactions to interactions in a completely novel technologically-mediated environment for these same tasks; the results of this study could then be compared with the predictions yielded by the benchmarks to draw conclusions regarding the viability of the benchmark approach in general.

By reducing the overall time required for the analysis, reducing the task time or number of environments examined could help to ameliorate one weakness in the Breakdown Analysis presented in this dissertation, by making it practical to examine more than four interactions in each environment. While the nonparametric statistical techniques used in this study are specifically designed for small sample sizes, increasing the number of data points in the statistical comparison from four in each environment to, for example, 12 or 15 would significantly strengthen the analysis.

Finally, it is important to point out that, regardless of how the methodology of Breakdown Analysis is modified to reduce the effort required, it will always be a relatively effort-intensive evaluative technique. While exact figures are not available, it is reasonable to estimate that evaluative techniques based on comparing user satisfaction or quality of work (see Chapter I) have AT:ST ratios somewhere between 5:1 and 20:1. Before deciding on Breakdown Analysis as a technique for evaluating the communicative efficacy of a technologically-mediated environment, it makes sense to consider the goals of the analysis. If the goal is merely to determine whether differences in communicative efficacy exist, then perhaps a comparison based on user satisfaction or quality of work would suffice. On the other hand, if the goal is to understand in detail how the technologies used in a technologically-mediated environment impinge on communicative interaction, what the limitations of those technologies are, and how one might ameliorate those limitations, then a Breakdown Analysis is appropriate.

7.3 Conclusion

As network connectivity and bandwidth continue to improve, it is likely that an increasing number of both personal and professional transactions will take place between participants interacting in some form of technologically-mediated environment. The goal of the study presented in this dissertation has been to articulate the pragmatic consequences of distributed interaction by exploring the functional differences between copresent interaction and interaction in two technologically-mediated environments representative of currently available technologies. By exposing the ways in which the communication environment impinges on participants' ability to efficiently and effectively accomplish their communicative goals, we can begin to understand the strengths and weaknesses of technologically-mediated interaction, and perhaps see how future designs might better support the collaborative activities of distributed participants.

The findings of this study are summarized in the following points:

1. The communicative efficacy of audio-only and audio-video environments was significantly lower than that of the copresent condition.

2. The higher incidence of communicative breakdown observed in distributed interactions was related to an overwhelming insensitivity to nonverbal displays like hand position, direction of gaze, and deictic gesture.

3. The video image of a conversational partner provided in the audio-video environment did not support access to that partner's nonverbal displays. There are profound pragmatic differences in the visual access afforded by a video image and that afforded by physical copresence.

In sum, the results of this study suggest that the intuitive notion that environments that provide a higher bandwidth connection between participants inherently support a higher communicative efficacy is overly simplistic — there is a great deal of difference between *technically* increasing the amount of communicative resources available in an environment and the *practical* utility of such upgrades to participants.

This observation bodes ill for the current crop of technologically-mediated environments, e.g. In Person, See-U-See-Me, and MMCC/VAT, available to the general public. At best, the video images provided in these environments will have no impact on the communicative efficacy of task-oriented interactions in these environments; participants will experience no less breakdown than if designers had simply left out the video connection. At worst, participants efforts to utilize the remote video image may actually lead to a higher incidence of breakdown, as participants divide their attention between the task representation and the remote video image.

A more general implication of this work is that it is unrealistic to place naive participants in a technologically-mediated environment and expect their interactions to be just as robust as if they were physically copresent. Even if it is possible to design technologically-mediated environments that overcome some of the limitations associated with remote video images, it is unlikely that interaction in any of these environments will ever be truly identical to copresent interaction. In particular, participants will need to learn to utilize whatever compensatory mechanisms the system provides as substitutes for copresent visual access. This implies that users might benefit greatly from some form of training to familiarize them with the limitations imposed by the environment, and actively teach them how to use the resources that *are* provided by the system to compensate for those limitations. Unfortunately, designers of commercial systems have placed little emphasis on training to date.

Finally, it is important to emphasize that the analysis presented in this dissertation has focused narrowly on the *functional* comparison of copresent and technologically interaction. In particular, there has been no effort to explore the social implications of technologically-mediated interaction. For instance, does interaction via a technologically-mediated environment have a lower social status that a face-to-face visit? Are commitments or decisions made during technologically-mediated interactions perceived as weaker than those made in copresent interaction? Is it possible to establish trust as effectively in technologically-mediated environments as when face-to-face? In what ways does this technology affect an individual's control over his or her personal privacy?

A growing body of evidence suggests that social issues like these may ultimately be *more* important determinants of whether a given technologically-mediated environment is actually used by participants in their everyday interactions than the functional differences explored in this work. For example, Hollan and Stornetta (1993) have argued that face-to-face interaction constitutes a unique "social glue" and that technologically-mediated interactions — no matter what the communicative efficacy of the environment — are fundamentally unable to generate the same level of trust and commitment as copresent interaction. On a more practical level, a least one study of how real world participants used a sophisticated audio-video environment found that some participants physically unplugged the system in order to regain control over their privacy and personal accessibility (Mantei, Baecker et al., 1991) .

These observations emphasize that it makes sense to think very carefully, not only about whether a given technologically-mediated environment supports the same communicative efficacy as copresent interaction, but also about how the environment will be integrated with the real world communicative activities of target users.

7.4 Notes

[1] In normal physical contexts, this criterion is met automatically when the previous two criteria are satisfied. However, this is not necessarily true of virtual reality systems (discussed below) which may or may not support peripheral vision.

[2] Another obvious difficulty in Clearboard is that text written by one participant on the virtual whiteboard is seen in reverse by the other participant.

[3] We can expect this to change, however, as software oriented towards multiple users becomes increasingly popular.

[4] The ratios presented in Table 7.1 are an estimate of the "best case" analysis time, where the analyst is familiar with the methodology, has applied it before, and is already skilled at audio-video transcription. Including the overhead associated with developing and applying a novel methodology, the AT:ST ratio actually required to perform the analysis presented in this dissertation was approximately 120:1.

[5] Of course, another way to reduce the total amount of videotape data would be to examine fewer pairs of participants in each environment. However, this would clearly compromise the statistical analysis and must be ruled out for that reason.

BIBLIOGRAPHY

Abel, M. J. (1993). Experiences in an Exploratory Distributed Organization. In R. M. Baecker (Ed.), Groupware and Computer-Supported Cooperative Work (pp. 792-802). San Mateo, CA: Morgan Kaufman Publishers, Inc.

Action Technologies (1989). The CO-ORDINATOR II.

Agre, P. (1993). The Symbolic Worldview: Reply to Vera and Simon. Cognitive Science, 17, 7-48.

Ahuja, S. R., Ensor, D. N., & Lucco, S. E. (1988). The Rapport multimedia conferencing system: A software overview. In Proceedings of 2nd IEEE Conference on Computer Workstations (pp. 52-58). Los Angeles: IEEE Press.

Anderson, A. H., Bader, M., Bard, E. G., Boyle, E., Doherty, G., Garrod, S., Isard, S., Kowto, J., McAllister, J., Miller, J., Sotillo, C., & Thompson, H. S. (1991). The HCRC Map Task Corpus. Language and Speech, 34, 351-366.

Apperley, M., & Masoodian, M. (1995). An experimental evaluation of video support for shared workspace interaction. In Proceedings of CHI'95 (pp. 306-307). Denver, CO: ACM Press.

Argyle, M. (1969). Social Interaction. London: Methuen.

Atkinson, J. M., & Heritage, J. (1984). Structures of Social Action. Cambridge: Cambridge Press.

Austin, J. L. (1962). How to Do Things with Words. Cambridge, MA: Harvard University Press.

Bartle, R. (1990). Interactive Multi-User Computer Games. Unpublished electronic document available at: ftp://ftp.ccs.neu.edu//pub/mud/docs/papers/ under file name mudreport.ps.gz.

Beach, W. (1990). Language as and in technology: Facillitating topic organization in a Videotex focus group meeting. In M. J. Medhurst, A. Gonzalez, & T. R. Peterson (Eds.), Communication and the Culture of Technology (pp. 34-50). Pullman: Washington State University Press.

Beach, W. A. (1993). Transitional regularities for 'casual' "Okay" useages. Journal of Pragmatics, 19, 325-352.

Benford, S., Bowers, J., Fahlen, L. E., Greenhalgh, C., & Snowdon, D. (1995). User Embodiment in Collaborative Environments. In Proceedings of CHI'95 (pp. 243-249). Denver, CO: ACM Press.

Bly, S. A., & Minneman, S. L. (1990). Commune: A Shared Drawing Surface. In Proceedings of Conference on Office Information Systems (pp. 184-192). Boston, MA: ACM Press.

Borning, A., & Travers, M. (1991). Two approaches to casual interaction over computer and video networks. In Proceedings of CHI'91 (pp. 13-19). New Orleans, LA: ACM Press.

Brachman, R. J., & Levesque, H. J. (Eds.). (1985). Readings in Knowledge Representation. Los Altos, CA: Morgan Kaufman.

Brown, J. S., Burton, R. R., & de Kleer, J. (1982). Pedagogical, natural language and knowledge engineering techniques in SOPHIE I, II, and III. In D. Sleeman & J. S. Brown (Eds.), Intelligent Tutoring Systems (pp. 227-282). London: Academic Press.

Button, G., & Casey, N. (1984). Generating topic: the use of topic initial elicitors. In M. J. Atkinson & J. Heritage (Eds.), Structures of Social Action (pp. 167-190). Cambridge: Cambridge Press.

Buxton, W. A., & Moran, T. (1990). EuroPARC's Integrated Interactive Intermedia facility (IIIF): Early Experiences. In Proceedings of IFIP Work Group8.4 Conference on Multi-user Interfaces and Applications (pp. 56-63). Heraklion, Crete: ACM Press.

Card, S., Moran, T., & Newell, A. (1983). The Psychology of Human-Computer Interaction. Hillsdale, NJ: Lawrence Erlbaum Associates.

Card, S., Robertson, G. G., & Mackinlay, J. D. (1991). The information visualizer: an information workspace. In Proceedings of CHI'91 (pp. 181-188). New Orleans, LA: ACM Press.

Champness, B. G. (1970). Mutual glance and the significance of look. Advancement of Science, 309-312.

Chapanis, A. (1975). Interactive Human Communication. Scientific American, 232, 34-42.

Clark, H. H., & Brennan, S. E. (1991). Grounding in Communication. In L. B. Resnick, J. M. Levine, & S. D. Teasley (Eds.), Perspectives on Socially Shared Cognition (pp. 127-150). Washington, DC: American Psychological Association.

Clark, H. H., & Schaefer, E. F. (1989). Contributing to Discourse. Cognitive Science, 13, 259-294.

Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a Collaborative Process. Cognition, 22, 1-39.

Comer, D., & Peterson, L. (1986). Conversation-based mail. ACM Transactions on Computer Systems, 4(4), 299-343.

Cook, M., & Lalljee, M. (1972). Verbal substitutes for verbal signals in interaction. Semiotica, 6, 212-221.

Covelli, S., & Murray, S. (1980). Accomplishing topic change. Anthropological Linguistics, 22, 382-389.

Curtis, P. (1992). MUDDING: Social Phenomena in Text-Based Virtual Realities. In Proceedings of 1992 Conference on Directions and Implications of Advanced Computing. Palo Alto, CA: ACM Press.

Doerry, E. (1994). Evaluating Foundations: A Critique of Traditional Conceptions of Knowing and Learning in Intelligent Tutoring. Technical Report: TR-94-24, University of Oregon, Eugene.

Douglas, S., & Doerry, E. (In preparation). Curricular Organization and Misinterpretation in dynamic simulations.

Douglas, S., & Liu, Z. Y. (1989). Generating Causal Explanation from a Cardio-Vascular Simulation. In Proceedings of 11th International Joint Conference on Artificial Intelligence. Los Angeles: Morgan Kaufman.

Douglas, S. A., & Doerry, E. (1994a). The Cardiovascular Construction Kit. Published on “The BioQUEST Library CD-ROM”, 3rd ed. Academic Software Development Group, University of Maryland, College Park, MD.

Douglas, S. A., & Doerry, E. (1994b). The Cardiovascular Construction Kit: A User’s Manual. Technical Report: TR-94-23, University of Oregon, Eugene.

Dourish, P., & Bly, S. (1993). Portholes: Supporting Awareness in a Distributed Work Group. In R. M. Baecker (Ed.), Groupware and Computer-Supported Cooperative Work (pp. 809-815). San Mateo, CA: Morgan Kaufman Publishers, Inc.

Downing, K. (1990). The use of teleology in the qualitative evaluation and explanation of circulatory systems. Unpublished Ph.D. dissertation, University of Oregon.

Dykstra-Erickson, E., Rudman, C., Hertz, R., Mithal, K., Schmidt, J., & Marshall, C. (1995). Supporting Adaptation to Multimedia Desktop Conferencing. In Proceedings of Proceedings of 15th International Symposium on Human Factors in Telecommunications (pp. 31-38). Melbourne, Australia: Humbolt Press.

Egan, K. (1989). Teaching as Story Telling. Chicago, IL: University of Chicago Press.

Ellis, C. A., Gibbs, S. J., & Rein, G. L. (1991). Groupware: some issues and experiences. Communications of the ACM, 34(1), 38-58.

Engelbart, D. C. (1975). NLS teleconferencing features: The Journal, and shared-screen teleporting. In Proceedings of Fall 1975 COMPCON (pp. 173-176). San Francisco, CA: IEEE Press.

Erickson, F. (1982). Money tree, lasagna bush, salt and pepper: Social construction of topical cohesion in a conversation among Italian-Americans. In D. Tannen (Ed.), Analyzing Discourse: Text and Talk (pp. 83-99). Washington, D. C.: Georgetown University Press.

Evard, R. (1993). Collaborative Networked Communication: MUDs as Systems Tools. In Proceedings of LISA 1993. Monterey, CA: ACM Press.

Fish, R. S., Kraut, R. E., & Chalfonte, B. L. (1990). The VideoWindow System in informal communication. In Proceedings of CSCW'90 (pp. 1-11). Los Angeles, CA: ACM Press.

Fox, B. A. (1993). The Human Tutorial Dialogue Project: Issues in the Design of Instructional Systems. Hillsdale, NJ: Lawrence Erlbaum.

Garfinkel, H. (1967). Studies in ethnomethodology. Englewood Cliffs, NJ: Prentice-Hall, Inc.

Gaver, W. W., Smets, G., & Overbeeke, K. (1995). A Virtual Window on Media Space. In Proceedings of CHI'95 (pp. 257-264). Denver, CO: ACM Press.

Goodwin, C. (1981). Conversational Organization: Interaction between a Speaker and Hearer. London: Academic Press.

Goodwin, C. (1984). Notes on story structure and the organization of participation. In M. J. Atkinson & J. Heritage (Eds.), Structures of Social Action (pp. 225-246). Cambridge: Cambridge Press.

Goodwin, C. (1986). Gesture as a resource for the organization of mutual orientation. Semiotica, 62(1/2), 29-49.

Greeno, J. (1993). Situativity and Symbols: Response to Vera and Simon. Cognitive Science, 17, 7-48.

Grudin, J. (1988). Why CSCW applications fail: Problems in the design and evaluation of organizational interfaces. In Proceedings of CSCW 1988 (pp. 85-93). Portland, OR: ACM Press.

Gust, P. (1989). SharedX: X in a distributed group work environment. In Proceedings of Second Annual X Conference. Boston, MA: MIT Press.

Hamm, M., & Adams, D. (1992). The Collaborative Dimensions of Learning. Norwood, NJ: Ablex Publishing Corp.

Harrison, S., & Minneman, S. (1990). The Media Space: a research project into the use of video as a design medium. In Proceedings of Conference on Participatory Design (pp. 51-58). Seattle, WA: ACM Press.

Heath, C. (1984). Talk and recipiency: sequential organization in speech and body movement. In M. J. Atkinson & J. Heritage (Eds.), Structures of Social Action (pp. 247-265). Cambridge: Cambridge University Press.

Heath, C. (1986). Body movement and speech in medical interactions. Cambridge: Cambridge University Press.

Heath, C., & Luff, P. (1992). Media space and communicative asymmetries: Preliminary observations of video-mediated interaction. Human Computer Interaction, 7, 315-346.

Heath, C., & Luff, P. (1993). Disembodied Contact: Communication through video in a multimedia office environment. In R. M. Baecker (Ed.), Groupware and Computer-Supported Cooperative Work (pp. 837-841). San Mateo, CA: Morgan Kaufman Publishers, Inc.

Hollan, J., & Stornetta, S. (1993). Beyond Being There. In R. M. Baecker (Ed.), Groupware and Computer-Supported Cooperative Work (pp. 842-848). San Mateo, CA: Morgan Kaufman Publishers, Inc.

Isaacs, E. A., & Clark, H. H. (1987). References in Conversation between Experts and Novices. Journal of Experimental Psychology, 116(1), 26-37.

Isaacs, E. A., Morris, T., Rodriguez, T. K., & Tang, J. C. (1995). A Comparison of Face-to-Face and Distributed Presentations. In Proceedings of CHI'95 (pp. 355-361). Denver, Colorado: ACM Press.

Ishii, H. (1990). TeamWorkStation: Towards a seamless shared workspace. In Proceedings of CSCW'90 (pp. 13-26). Los Angeles, CA: ACM Press.

Ishii, H., & Kobayashi, M. (1993). ClearBoard: A Seamless Medium for Shared Drawing and Conversation with Eye Contact. In R. M. Baecker (Ed.), Groupware and Computer-Supported Cooperative Work (pp. 829-836). San Mateo, CA: Morgan Kaufman Publishers, Inc.

Jefferson, G. (1984). On the organization of laughter in talk about troubles. In M. J. Atkinson & J. Heritage (Eds.), Structures of Social Action (pp. 346-369). Cambridge: Cambridge Press.

Jefferson, G. (1988). Notes on a possible metric which provides for a 'standard maximum' silence of approximately one second in conversation. In D. Roger & P. Bull (Eds.), Conversation: An interdisciplinary perspective (pp. 121-143). Clevedon, UK: Multilingual Matters.

Jefferson, G. (Ed.). (1992). Lectures on Conversation. Cambridge, MA: Blackwell Publishers.

Jefferson, G. (1993). Caveat speaker: Preliminary notes on recipient topic-shift implicature. ROSLI, 26, 1-30.

Jordan, B., & Henderson, A. (1995). Interaction Analysis: Foundations and Practice. Journal of the Learning Sciences, 4, 39-103.

Kaplan, S. (1990). COED: A conversation-oriented tool for coordinated design work. In Proceedings of IFIP Int. Workshop on Human Factors in Information Systems. Boston, MA: ACM Press.

Kendon, A. (1967). Some functions of gaze direction in social interaction. Acta Psychologica, 26, 1-47.

Kiesler, S., Siegel, J., & McGuire, T. (1991). Social Psychological Aspects of Computer-Mediated Communication. In C. Dunlop & R. Kling (Eds.), Computerization and Controversy (pp. 330-349). Boston, MA: Academic Press.

Kimball, R. (1982). A self-improving tutor for symbolic integration. In D. Sleeman & J. S. Brown (Eds.), Intelligent Tutoring Systems (pp. 283-308). London: Academic Press.

Knister, M. J., & Prakash, A. (1990). DistEdit: A distributed toolkit for supporting multiple group editors. In Proceedings of CSCW'90 (pp. 343-355). Los Angeles, CA: ACM Press.

Kraut, R. E., Fish, R. S., Root, R. W., & Chalfonte, B. L. (1993). Informal Communication in Organizations: Form, Function, and Technology. In R. M. Baecker (Ed.), Groupware and Computer-Supported Cooperative Work (pp. 287-314). San Mateo, CA: Morgan Kaufman Publishers, Inc.

Laird, J. E., Newell, A., & Rosenbloom, P. S. (1987). Soar: An architecture for general intelligence. Artificial Intelligence, 33, 1-64.

Lauwers, J. C., & Lantz, L. A. (1990). Collaboration Awareness in Support of Collaboration Transparency: Requirements for the next generation of shared window systems. In Proceedings of CHI'90 (pp. 303-311). Seattle, WA: ACM Press.

Lave, J., & Wenger, E. (1991). Situated Learning: Legitimate Peripheral Participation. Cambridge, MA.: Cambridge University Press.

Lawrence Livermore Labs (1992). The VAT, NV, and MMCC protocols. Published as public domain software, Lawrence Livermore Labs, Oakland, CA.

Lewis, B. J., Koved, L., & Ling, D. T. (1991). Dialogue Structures for Virtual Worlds. In Proceedings of CHI'91 (pp. 131-136). New Orleans, LA: ACM Press.

Mackay, W. E. (1989). EVA: An experimental video annotator for symbolic analysis of video data. SIGCHI Bulletin, 21, 68-71.

Mackinlay, J. D., Robertson, G. G., & Card, S. K. (1991). Rapid controlled movement through virtual 3D workspaces. In Proceedings of CHI'91 (pp. 455-456). New Orleans, LA: ACM Press.

Mantei, M. M., Baecker, R. M., Sellen, A. J., Buxton, W. A., & Milligan, T. (1991). Experiences in the Use of a Media Space. In Proceedings of CHI'91 (pp. 203-208). New Orleans, LA: ACM Press.

McCarthy, J. C., Miles, V. C., & Monk, A. F. (1991). An experimental study of common ground in text-based communication. In Proceedings of CHI'91 (pp. 209-215). New Orleans, LA: ACM Press.

Minneman, S. L., & Bly, S. A. (1991). Managing à trois: a study of a multi-user drawing tool in distributed design work. In Proceedings of CHI'91 (pp. 217-224). New Orleans, LA: ACM Press.

Newell, A. (1980). Physical Symbol Systems. Cognitive Science, 4, 135-183.

Oikarinen, J. (1988). Internet Relay Chat. Published as public domain software, University of Oulu, Oulu, Finland.

Okada, K.-i., Maeda, F., Ichikawaa, Y., & Matsushita, Y. (1994). Multiparty Videoconferencing at Virtual Social Distance: MAJIC Design. In Proceedings of CSCW'94 (pp. 385-395). Chapel Hill, NC: ACM Press.

Olson, J. S., Olson, G. M., & Meader, D. K. (1995). What Mix of Video and Audio is Useful for Small Groups Doing Remote Real-time Design Work? In Proceedings of CHI'95 (pp. 362-368). Denver, Colorado: ACM Press.

Oviatt, S. L., & Cohen, P. R. (1989). The effects of interaction of spoken discourse. In Proceedings of 27th Meeting of the Association of Computational Linguistics (pp. 126-134).

Parsons, T. (1937). The Structure of Social Action. New York: McGraw-Hill.

Pausch, R. (1991). Virtual Reality on Five Dollars a Day. In Proceedings of CHI'91 (pp. 265-270). New Orleans, LA: ACM Press.

Pomerantz, A. M. (1975). Second Assessments: A Study of Some Features of Agreements/Diagreements. Unpublished Ph.D. dissertation, University of California at Irvine.

Pomerantz, A. M. (1978). Compliment responses: notes on the co-operation of multiple constraints. In J. Schenkein (Ed.), Studies in the Organization of Conversational Interaction. (pp. 79-112). London: Academic Press.

Psathas, G., & Anderson, T. (1990). The Practices of Transcription in Conversation Analysis. Semiotica, 78, 75-99.

Reddy, M. (1979). The Conduit Metaphor: A Case of Frame Conflict in Language about Language. In A. Ortony (Ed.), Metaphor and Thought (pp. 66-102). Cambridge: Cambridge University Press.

Reid, E. M. (1992, Winter). Electropolis: Communication and Community on the Internet Relay Chat. Intertek, p. 22-30.

Robertson, G. G., Mackinlay, J. D., & Card, S. K. (1991). Cone Trees: Animated 3d visualizations of hirarchical information. In Proceedings of CHI'91 (pp. 189-194). New Orleans, LA: ACM Press.

Root, R. W. (1988). Design of a multi-media vehicle for social browsing. In Proceedings of CSCW'88 (pp. 25-38). Portland, OR: ACM Press.

Sacks, H. (1992a). Bases for ‘interruption’. In G. Jefferson (Ed.), Lectures on Conversation (pp. 641-646). Cambridge, MA: Blackwell Publishers.

Sacks, H. (1992b). The ‘one-at-a-time’ rule; Violations; Complaints; Gossip. In G. Jefferson (Ed.), Lectures on Conversation (pp. 633-640). Cambridge, MA: Blackwell Publishers.

Sacks, H. (1992c). Sequencing: Utterances, Jokes, and Questions. In G. Jefferson (Ed.), Lectures on Conversation (pp. 95-103). Cambridge, MA: Blackwell Publishers.

Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for organization of turn-taking in conversation. Language, 50(4), 696-735.

Sanderson, P. M., & Fisher, C. (1994). Exploratory Sequential Data Analysis: Foundations. Human-Computer Interaction, 9(3,4), 251-318.

Schegloff, E. (1987). Recycled turn beginnings: A precise repair mechanism in conversation’s turn-taking organization. In G. Button & J. R. E. Lee (Eds.), Talk and social organization (pp. 137-168). Clevedon, UK: Multilingual Matters.

Schegloff, E., Jefferson, G., & Sacks, H. (1977). The preference for self-correction in the organization of repair in conversation. Language, 53, 361-382.

Schegloff, E. A. (1972). Sequencing in Conversational Openings. In J. J. Gumperz & D. Hymes (Eds.), Directions in Sociolinguistics (pp. 346-380). New York, NY: Holt, Rineman and Winston.

Schegloff, E. A. (1979). Identification and recognition in telephone conversation openings. In G. Psathas (Ed.), Everyday Language: Studies in ethnomethodology (pp. 23-78). New York, NY: Irvington Publishers.

Schegloff, E. A. (1984). On some gestures’ relation to talk. In M. J. Atkinson & J. Heritage (Eds.), Structures of Social Action (pp. 266-296). Cambridge: Cambridge Press.

Schegloff, E. A., & Sacks, H. (1973). Opening up Closings. Semiotica, 7, 289-327.

Searle, J. R. (1979). Expression and Meaning: Studies in the Theory of Speech Acts. Cambridge: Cambridge University Press.

Shepherd, A., Mayer, N., & Kuchinsky, A. (1990). Strudel - An Extensible Electronic Conversation Toolkit. In Proceedings of CSCW'90 (pp. 93-104). Los Angeles, CA: ACM Press.

Short, J., Williams, E., & Christie, B. (1976). The Social Psychology of Telecommunications. New York: John Wiley and Sons.

Siegel, S., & Castellan, N. J. (1988). Nonparametric statistics for the behavioral sciences. New York: McGraw-Hill.

Slavin, R. (1983). Cooperative Learning. New York: Longman Publishers.

Stefik, M., Foster, G., Bobrow, D., Kahn, K., Lanning, S., & Suchman, L. (1988). Beyond the chalkboard: Computer-supported collaboration and problem solving in meetings. In I. Greif (Ed.), Computer-Supported Cooperative Work: A Book of Readings (pp. 335-366). San Mateo, CA: Morgan Kaufman.

Suchman, L. (1987). Plans and Situated Actions: The Problem of Human-Machine Communication. Cambridge, England: Cambridge University Press.

Suchman, L. (1993). Response to Vera and Simon's Situated Action: A Symbolic Interpretation. Cognitive Science, 17, 7-48.

Suchman, L., & Trigg, R. (1991). Understanding practice: Video as a medium for reflection and design. In J. Greenbaum & M. Kyng (Eds.), Design at work (pp. 65-89). Hillsdale, NJ: Lawrence Erlbaum Associates.

Tang, J. (1991). Findings from observational studies of collaborative work. International Journal of Man-Machine Studies, 34, 134-160.

Tang, J. C., & Isaacs, E. (1992). Why do users like video? Studies of multi-media supported collaboration. In Proceedings of CSCW'92 (pp. 163-196). Toronto: ACM Press.

Tang, J. C., Isaacs, E. A., & Rua, M. (1994). Supporting Distributed Groups with a Montage of Lightweight Interactions. In Proceedings of CSCW'94 (pp. 385-395). Chapel Hill, NC: ACM Press.

Tang, J. C., & Minneman, S. L. (1990). VideoDraw: A video interface for collaborative drawing. In Proceedings of CHI'90 (pp. 313-320). Seatle, WA: ACM Press.

Tang, J. C., & Minneman, S. L. (1991). VideoWhiteboard: Video shadows to support remote collaboration. In Proceedings of CHI'91 (pp. 315-322). New Orleans, LA: ACM Press.

Tatar, D. (1989). Using Video-Based Observation to Shape the Design of a New Technology. SIGCHI Bulletin, 21(2), 108-111.

Teasley, S. D., Olson, J. S., & Meader, D. K. (in preparation). Toward the Task/Technology Fit for Group Work, An Analysis of Task Characteristics. .

Tukey, J. W. (1977). Exploratory Data Analysis. Reading, MA: Addison-Wesley.

VanLehn, K. (1990). Mind Bugs. Cambridge, MA: MIT Press.

Vera, A. A., & Simon, H. A. (1993). Situated Action: A Symbolic Interpretation. Cognitive Science, 17, 7-48.

Weiser, M. (1991). The computer for the 21st century. Scientific American, 265(3), 94-104.

West, C., & Garcia, A. (1988). Conversational shift work: A study of topical transitions between women and men. Social Problems, 35, 551-575.

Whalen, J. (1995). A Technology of Order Production: Computer-Aided Dispatch in Public Safety Communications. In P. ten Have & G. Psathas (Eds.), Situated Order: Studies in the social organization of talk and embodied action. Washington, D.C.: University Press.

Wilkes-Gibbs, D., & Clark, H. H. (1992). Coordinating Beliefs in Conversation. Journal of Memory and Language, 31, 183-194.

Winograd, T., & Flores, F. (1986). Understanding Computers and Cognition. Norwood, NJ: Ablex Corporation.

Wooffitt, R. (1990). On the Analysis of Interaction. In P. Luff, N. Gilbert, & D. Frohlich (Eds.), Computers and Conversation (pp. 7-38). London: Academic Press Ltd.

APPENDIX A

THE CARDIOVASCULAR CONSTRUCTION KIT (CVCK)

The Cardiovascular Construction Kit was designed and implemented over the course of about five years, starting in 1987, as the main focus of a project sponsored by a FIPSE grant (Douglas & Liu, 1989; Downing, 1990) . The goal of this research was to explore the mental models of cardiovascular dynamics used by naive learners, and to design an Intelligent Tutoring System to support the formation of "correct" conceptions in this domain.

To help readers unfamiliar with this work to understand the references to it throughout this dissertation, this appendix provides a brief introduction.

A.1 The Cardiovascular Construction Kit

The Cardiovascular Construction Kit is a computer simulation of a very unusual experimental laboratory — so unusual that it could never exist in any real sense. In this laboratory, students are able to piece together arbitrarily complex cardiovascular systems using a pre-defined palette of components, "run" the resulting construction to observe its dynamic behavior, and measure and compare certain simulation parameters in order to reach general conclusions about cardiovascular behavior. Compared to other tutoring systems, which typically provide the student with a sequence of predetermined problems (Brown, Burton, & de Kleer, 1982; Kimball, 1982; VanLehn, 1990) to work on, CVCK allows the learner to explore the problem space freely. While this feature has been found to be of marginal utility for novice users, who have little idea of how to structure their exploration, a central goal of the project was to support more advanced learners as well.

As users become more sophisticated and start to form hypotheses about the relationships between structural characteristics of a cardiovascular system and the system's behavior, they can immediately test such hypotheses by modifying the system and running it again to observe the effect of the changes.

The best way to describe the CVCK is by considering the way in which it is actually used by the learner. Interaction with CVCK can be decomposed into three distinct types of activity: construction, measurement, and observation.

Construction. Though it is possible for the user to load pre-existing labs for further experimentation, the most common way to initiate a laboratory session is to construct a cardiovascular system "from scratch." Users are presented with a blank work area, into which they can drag components from an iconic palette. These components are connected by simply arranging them adjacent to one another in the desired configuration. At any point, the user can double-click on a component and adjust component specific parameters. For instance, the diameter and elasticity parameters of the "vessel" component can be adjusted in this way.

Measurement. All components have attachment points for gauges. Like components themselves, gauges may be dragged into the workspace from the palette, attached to any attachment point, and set to measure one of several pre-defined quantities (e.g. pressure, flow, oxygen concentration). Gauges may be "opened" into a two-dimensional graph showing the behavior of the measured quantity over time. This display is updated dynamically as the simulation runs. Thus, the gauges show the recent history of behavior of the measured quantity, at that point in the construction. For example, a gauge might be set to document the changing pressure value inside the ventricle. Gauges may be moved about and juxtaposed for purposes of comparison, remaining visually attached to their attachment points by a thin line. Importantly, all measurements and settings in the CVCK are *qualitative* instead of quantitative, reflecting an experimental hypothesis that learners reason qualitatively before they reason quantitatively.

Observation. After constructing the desired cardiovascular configuration, the user sets the simulation in motion using a "control panel," which is similar to the controls found on any VCR. The user may run the simulation, pause it, or slow down the action, as well as setting the number of "heartbeats" that the simulation is to run. While the simulation is running, the certain components are animated (e.g. the ventricle expands and contracts as it beats), the gauges are updated continuously, and small arrows appear in each component to indicate the instantaneous direction of blood flow. Based on observations of run-time behavior and subsequent analysis of values documented by gauges attached to the construction, the student is asked to make generalizations about cardiovascular physics. For instance, we hoped that students would discover abstract causal relationships (e.g. "pressure difference causes flow").

Ideally, experimentation with one construction would expose further issues to be explored, motivating the user to iteratively return to the previous two steps to modify the construction, attach other gauges, and run the simulation again.

Figure A.1 shows the CVCK screen with a completed construction and attached gauges.

A.2 Discussion

The CVCK system has been refined and used extensively in the biology labs at the University of Oregon. Recently, the CVCK was published as part of a collection of biology-related software through the Bioquest project (Douglas & Doerry, 1994a) .

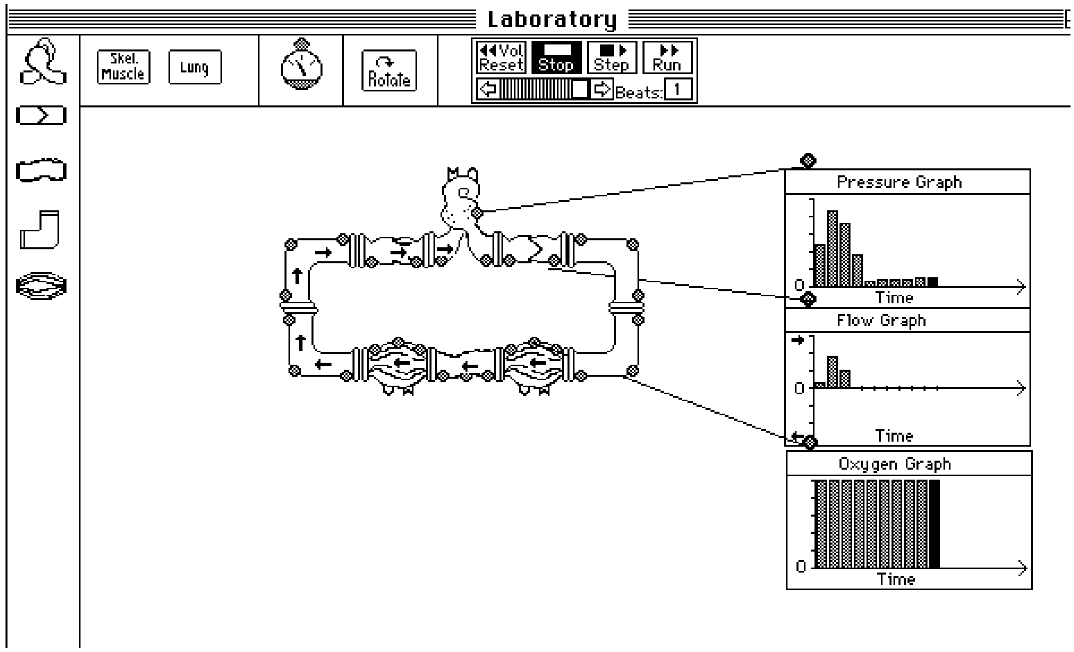


Fig. A.1: A simple construction in CVCK, with gauges attached and showing recent parameter histories.

APPENDIX B

LABORATORY SETUP AND EQUIPMENT

A general description of the data collection approach was presented in Chapter III. While that description is adequate for understanding the research presented and evaluating its merit, it does not provide a detailed account of how data collection was actually accomplished. It is the purpose of this appendix to provide this account.

In the following sections, two aspects of the experimental design and execution are documented in detail:

1. Technical Specifications. A detailed description of equipment used and the wiring schematics implemented to create the three electronically mediated communication environments explored in this research.

2. Actual Environment. This section gives a detailed pictorial and verbal account of the environment created, documenting camera angles, placement of computer and television monitors and so on.

Together, these sections provide a solid basis for understanding and replicating the results reported in this dissertation.

B.1 Technical Specifications

The technical obstacles to be overcome in collecting the data for this research were numerous and non-trivial, especially for the two distributed communication environments. This section documents the equipment and wiring arrangements used for each of the three environments.

B.1.1 Face-to-Face Interaction

This communication environment was clearly the easiest of the three to implement and capture on videotape. Since the participants are copresent and seated in front of the same computer screen, they automatically have access to a shared audio, shared video and shared workspace environment. The only technical challenges center around finding an effective way to record the interaction.

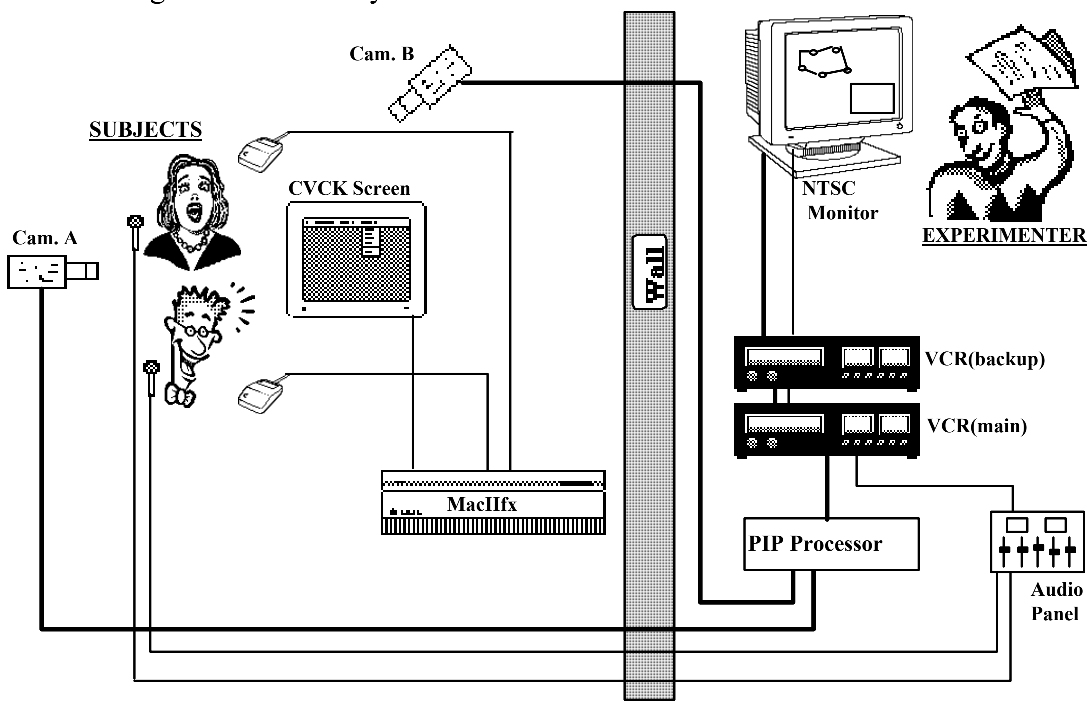


Figure B.1: Schematic of recording arrangement for face-to-face sessions.

Figure B.1 illustrates the arrangement used. Participants were placed in a room together with the computer (MacIIfx) running the CVCK simulation. Two cameras were used to capture the interaction. Camera A was mounted high and behind participants, shooting over and between their heads to record participants' actions within the electronic workspace and the dynamic behavior of the simulation. It also captured deictic gestures to objects on the screen made by participants during the interaction. Camera B was mounted high slightly behind and to the left of the workspace screen, providing a face-on shot of both participants. Importantly, this image was also framed to capture arbitrary gestures made by participants and to show the mouse controlled by each participant. This latter cue made it possible to determine which participant was controlling the screen cursor at any given moment.

Participants were each fitted with a high quality lapel microphone to capture all audio. The microphones were sufficiently sensitive that they were able to capture not only the speech of participants, but also any sounds (e.g. an error tone) made by the machine, and the click of the mouse. This latter cue was crucial in determining precisely when user actions were initiated.

Each participant was provided with a mouse; however, both mice controlled a single cursor on the screen. This meant that participants had to take turns using the mouse to work effectively.

Finally, the two participants shared a lab manual, containing the instructions for the various tasks attempted, and served to record answer to written questions about the simulation's behavior.

The images from these two cameras were wired through the wall to an adjacent control room, where the workspace image was inset into the face-on image using a picture-in-picture (PIP) video processor. Though this technique inevitably results in some data loss (as part of the main image is obscured by the inset), the framing of the face-on image was arranged to provide a non-critical space (i.e. the wall behind and above participants) over which to place the inset. The combined image was sent to a VCR to be recorded. The microphone output was first sent to an audio patch panel for amplification and then sent through the wall to the VCR as well. Both audio and visual images were continuously monitored on a television set in the control room. Since the videotape data is an irreplaceable resource for this work, the audio and video streams were also passed on to a second VCR to simultaneously create a backup copy.

Equipment Specifications:

VCR (main): Panasonic PV-4960 stereo
VCR (backup): Panasonic PV-S4864 stereo
PIP processor: Multivision model 1.1
Microphones: Realistic 33-1063 lapel
Audio Panel: Realistic 32-1100A
Camera A: Panasonic WV-3260 pro
Camera B: Panasonic PV-S350D
Control Room Monitor: Sony KV-27TS30 stereo television monitor.
Computer: MacIIfx with 13" color monitor set to black&white mode.
Mice: Two standard Macintosh mice connected to ADB bus; mouse driver set to "medium".
Cabling: Video passed through coax with BNC or RCA ends. Audio passed through standard two-conductor audio cable with RCA ends.

B.1.2 Audio-Only and Audio-Video Environments

From a technical standpoint, the arrangements for the two distributed communication environments were identical. In both cases, it was important to record both audio and video of each participant; the only difference was that, in the audio-only environment, the monitors provided for displaying the image of the remote participant were not turned on.

Implementing the two distributed environments presented several distinct and very challenging technical obstacles. It was necessary to somehow implement a synchronous shared view of the workspace in which both participants could move the cursor to point, gesture or initiate actions in the simulator. At the same time, participants had to be provided with a high-quality audio link and, in the audio-video scenario, with a crisp video image of the remote participant. Finally, all three images — the evolving workspace, and images of each participant — had to somehow be captured on videotape.

To avoid presenting nightmarish wiring schematics, I address each of these information channels separately.

B.1.3 Providing a Shared Workspace

The problem of providing shared access to an electronic (virtual) workspace is one of those problems that turns out to be much more challenging than it appears, especially when the shared application is graphics oriented. Since graphical entities require relatively high bandwidth to transmit, implementations using standard network links inevitably result in slight updating delays at the remote sites. For instance, we considered using a popular networking application known as Timbuktu™ to implement sharing of the CVCK workspace. When graphical entities (i.e. about anything in the CVCK) were moved about, the remote computer exhibited a “jerking” or “jumping” behavior. In a rapid series of action, whole actions by the local user might be lost, swallowed by the processing and transmission delays. This is not acceptable for this research. Since users are naive learners trying to make sense of an unfamiliar context, it is crucial that they both see precisely the same behavior. Moreover, they must also see shared behavior at precisely the same time -- any lag in transmission will throw off the mutually constitutive synchrony of talk and action.

One solution is to incorporate specialized high-speed links. For instance, a recent project (Dykstra-Erickson, Rudman et al., 1995) connected distributed participants using advanced fiber-optic technology. While this would be ideal, the budget constraints of this project dictated a more economical low-tech solution. Since participants in this series of protocols were in adjacent rooms rather than miles apart, we decided to simply connect two monitors to the same computer. As it turns out, this is not just a matter of buying or building a splitter cable to plug into the MacII video card. Since Mac monitors are terminated (rather than pass-through) devices, each monitor would sink (ground) the signal. At best, this results in a half-brightness image on each monitor; at worst it overloads the video card. The solution was to purchase a video distribution amplifier designed (apparently) for multimedia presentation or training systems. At a cost of $350 (device plus a custom 15 ft. video cable), the video distribution problem was solved. The resulting wiring schematic is shown in Figure B.2.

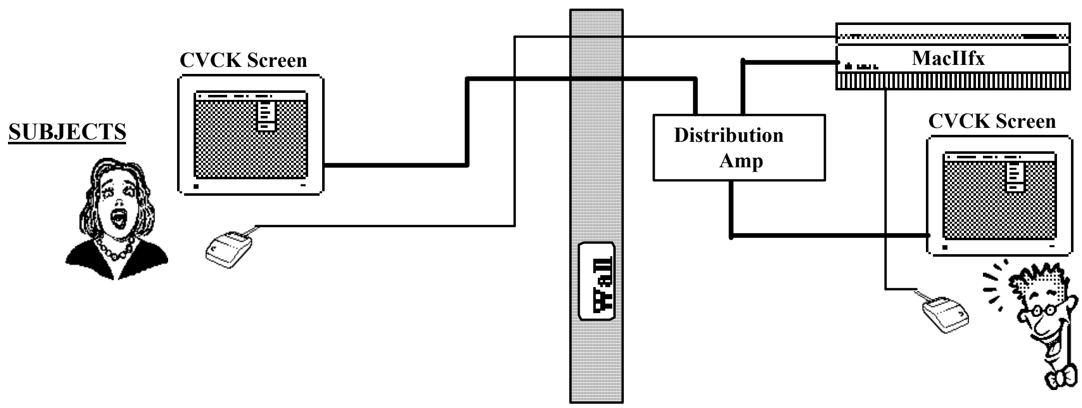


Figure B.2: Scenario 2&3 Shared Workspace Implementation. Dark lines are video.

Output from the MacII video card is sent to the distribution amp, and from there to the two monitors. In this way, actions on the local screen are instantaneously echoed on the remote screen.

Managing remote mouse input was substantially easier: A standard Macintosh ADB cable was cut and extended to approximately 17 ft., splicing in standard four-conductor telephone wire.

Equipment Specifications:

Distribution Amp: Extron Mac2-DA2 (contact: 1-800-882-7117)
Video Cable: Extron custom built 15 ft. Mac video

B.1.4 Distributing and Recording Audio

As it turned out, providing high quality audio for both participants and the audio/video record was one of the most frustrating obstacles to overcome. The combination of a demanding distribution schema and largely unshielded (i.e. standard) audio patch cables led to persistent problems with interference feedback and poor sound quality. In the end, the solution was to build coaxial cables with RCA connectors from the longer legs of the distribution schema. The schematic for the sound circuit is shown in Figure B.3.

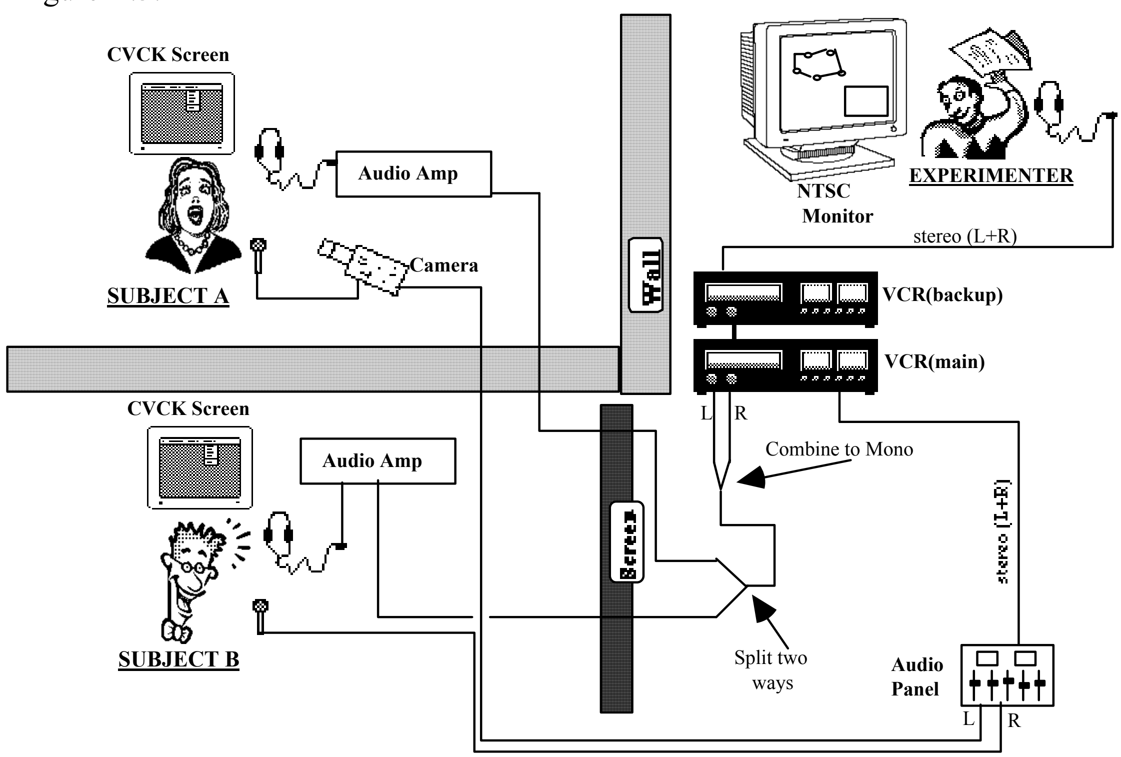


Figure B.3: Scenario 2&3 Audio Schematic. Camera included only because it was used as audio pre-amp.

Both subjects were fitted with lapel microphones as before. Since subject A was much farther away from the recording equipment than subject B, this sound channel was first passed through a pre-amplifier (the one in one of the video cameras was used) before being sent on to the audio patch panel. At the patch panel, the sound is mixed appropriately and sent on to the VCRs and, eventually, to the monitoring station. Note that, at this point, sound from each participant's microphone is placed on a separate channel. This makes it much easier for the analyst to later keep track of who is making what sound (including non-speech sound like mouse clicks, sighs, and so on). However, it is somewhat annoying for participants to hear themselves in one ear and their partner in the other. For this reason, sound distributed to participants is first combined into monaural sound and then split for delivery to individual participants. This is done using simple audio Y-splitter cables. Finally, the sound delivered to participants is amplified and heard using headphones. The headphones are important, as they prevent feedback from developing.

Equipment Specifications (additional):

Camera used for pre-amp: Panasonic WV-3260
Headphones: Tandy NOVA-35
Audio amp 1: Panasonic NV-8500 VCR (only used amp. circuit)
Audio amp 2: Panasonic RX-DS620 Portable stereo
Cables: 75 ohm coax and standard audio patch cords

B.1.5 Distributing and Recording Video

The only real challenge to distributing and recording the appropriate video images lay in finding a way to record three images on a single videotape. Although one could record the images on two, or even three, separate videotapes, this would lead to nearly insurmountable synchronization problem during the transcription process. After investigating numerous special purpose devices for placing of up to four video inputs within a single NTSC video frame (ranging from $800 to $4500), a more economical alternative was discovered: Use two PIPs connected in series. The first one insets one image, the second one insets the other into the output from the first. The resulting schematic is shown is Figure B.4.

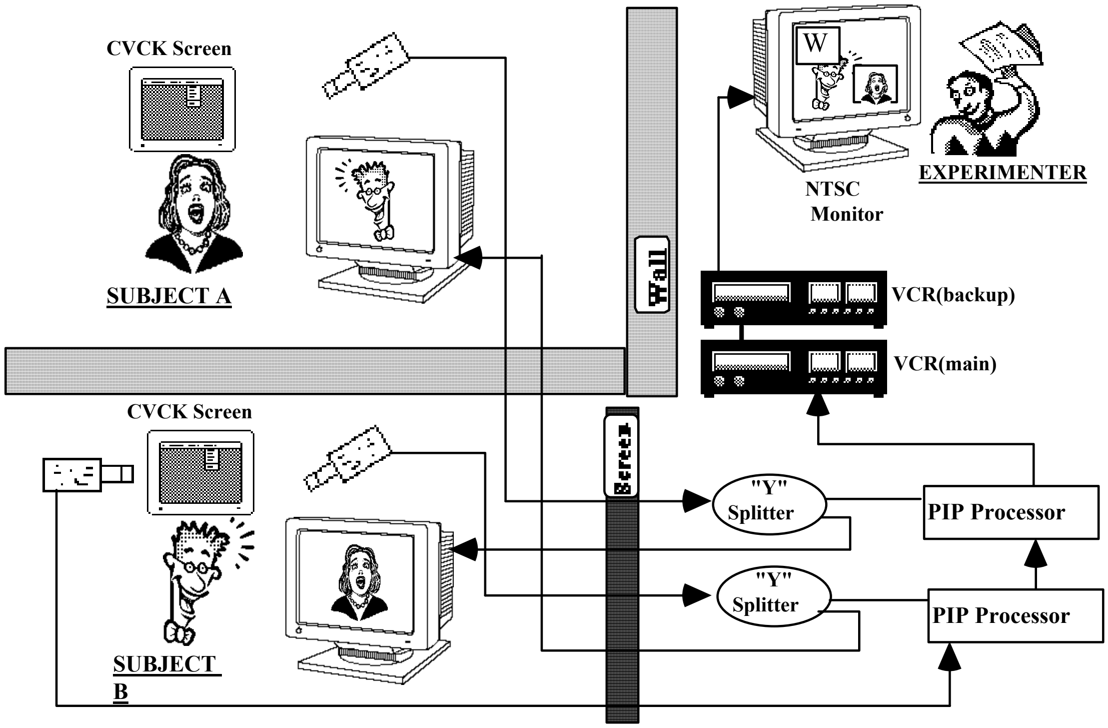


Figure B.4: Scenario 2&3 video schematic. The “W” in the experimenters monitor represents the inset picture of the workspace.

The figure indicates an interesting feature of the arrangement: While the experimenter must capture all three images, the participants must only be shown one, namely, the image of the other participant. This is accomplished by merely splitting the video signal before it enters the PIP processors and sending it to the appropriate participant’s monitor.

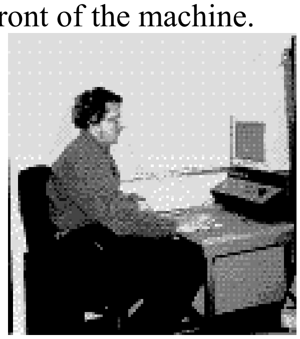
Equipment Specifications (additional):

Second PIP processor: RocTec RN1812 PIPview
Additional Camera: Ricoh R-86S
Participants Monitor 1: Sony KV-27TS30 color monitor
Participants Monitor 2: Tektronix 69M01 color monitor
Experimenter Monitor: Tektronix 650HR-1 color monitor

B.2 Actual Communication Environment

Even after describing the communication environments created for this research from a technical standpoint, it is difficult to visualize the actual environment created, especially for the distributed scenarios. This section briefly describes the physical environments created.

The protocols were confined to two adjacent rooms in our lab, which we will call the *main lab* and the *video lab*. In the first (face-to-face) communication scenario, the two participants were placed in the video lab along with the cameras and the audio panel; the main lab served as a control center and monitoring station, containing all of the video processors and recording equipment. In this scenario, participants were simply seated together in front of the machine.



(a)

(b)

Figure B.5: Main lab (a) participant work area and (b) control area.

In the second and third (distributed) communication environments, the main lab was partitioned into two sections using a portable room divider. The control and monitoring area remained in one partition; a participant work area was created in the other. The two partitions are depicted in Figure B.5.

The other participant was placed in the video lab. In each case, a camera was set to record the participant's face and upper body, as well as the mouse movement surface. This latter cue is important in allowing the analyst to retrospectively determine which participant is controlling the screen cursor from moment to moment. In the video lab, an additional camera was set to record the computer screen. The camera arrangement are shown in Figure B.6.

(a)

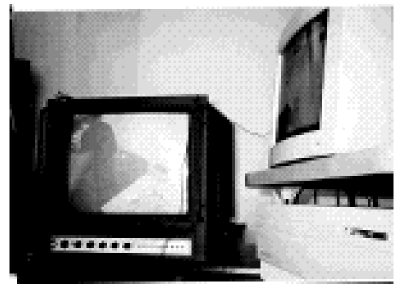
(b)

Figure B.6: Camera arrangement in (a) main lab and (b) video lab.

In the third (audio/video) sequence of protocols, participants had access to the shared computer screen, a shared audio channel, and a video image of the other participant. The remote video image was displayed on a separate monitor placed at approximately a 70 degree angle to the computer screen (shared workspace). The goal here was to force participants to turn their heads slightly when looking at the remote image. In this way, the analyst is able to determine when participants are looking at the workspace versus when they are looking at the remote participant.

The importance of mutual eye gaze has been extensively documented (Short, Williams et al., 1976) . In an ideal arrangement, the camera recording participant A would somehow be mounted behind the screen of A's remote monitor. In this way, when A looks at the remote image of B, it would appear to B that A is looking directly at him or her. In fact, Buxton (1990) has experimented with such technology. For these protocols, the remote cameras were merely placed close to the remote screen; while mutual eye gaze is not supported, participants do appear (to the other participant) to turn and look roughly at the other participant. The arrangement of the monitors in each work area is shown in figure B.7.

(a)



(b)

Figure B.7: Monitor arrangement in (a) main lab and (b) video lab.

B.4 Summary

Executing the research described in this dissertation required a significant investment in time and equipment as well as considerable technical ingenuity. Recruiting participants was an arduous and frustrating task; college students are not easily convinced to find time to participate in research, even when they are paid for it. Considering the various obstacles, the large amount of audio/video equipment required, and the very tight budget constraints, the data collection phase of this research went amazingly well.

APPENDIX C

DOCUMENTS

This appendix presents two important documents that were used in collecting the data for the study presented in this dissertaton. As described in Chapter III, pairs of participants for the study were selected from a pool of applicants by evaluating their answers to questions posed in a brief questionairre aimed at exposing participants' educational level and previous experience with computers. A second important document used in this study is the laboratory manual given to participants describing the task they were to perform using the CVCK simulator. Both documents are reproduced in the following pages.

Evaluating the Communicative Efficacy of Technologically-mediated environments

Questionnaire

Thank you for responding to my advertisement requesting volunteers for my study of collaborative interaction in distributed environments. I would like you to fill out the following form, which I will use to screen participants for the study. **If you have a partner that you would like to work with in the study, please fill out this questionnaire together.** Otherwise, simply leave the "participant #2" information blank and I will assign you a partner. I will be accepting 32 participants for the study, and if you are accepted, I will contact you before November 15, 1994.

In addition to screening participants, the information requested in this questionnaire will help me to evaluate the results of my study. **This study is anonymous!** The first page of this questionnaire (containing your personal information) will be destroyed immediately after you participate; your name(s) and personal information will not be attached to the data in any way. In addition, all data, including this questionnaire, will be treated with the strictest confidentiality. If you decide not to participate, or are not selected for the study, this questionnaire will be destroyed immediately.

Please use BLOCK LETTERS to fill out this questionnaire.

Personal Information

Participant #1:

Name:____

Sex:___________________ Date of Birth:_____________________

Contact Phone Number: ____

Participant #2:

Name:____

Sex:___________________ Date of Birth:_____________________

Contact Phone Number: ________________________________

Education:

Topics	Participant #1	Participant #2
Last Degree Received:		
Degree you are currently pursuing:		
Major:		

Background:

Please use the following scale to answer the next group of questions:

1 — Very familiar, use it all the time **2** — Quite familiar, I've used it often.
3 — Basic Knowledge, used it once or twice **4** — Heard of it, but never used it
5 — Unfamiliar, never heard of it

Topics	Participant #1	Participants #2
Electronic mail		
Video-telephone		
Conference Calling		
Computers in general		
Electronic Word Processing		
Macintosh computers		

Any computer simulation (video games don't count)		
The Internet		
MUDS or Chat Boards		

Cardiovascular Construction Kit Laboratory Workbook

Introductory Notes

The following instructions are purposefully vague! The idea is that the system should be simple and intuitive enough to use without much explanation. Furthermore, my interest is not so much in what or how much you get done; I am interested in how you and your partner collaborate as you proceed.

Mark your answers to the questions directly in the lab workbook.

Do your best and HAVE FUN!

Exercise 1: Constructing a Cardio-Vascular system

The first exercise requires you to construct a simple cardiovascular system, and to observe it running. The system you will construct has a very simple "heart", namely a ventricle plus two one-way valves.
(*Biology note*: Of course, the heart of a human would have two such ventricle-valve structures.)

1. Use the palette of components on the left side of the workspace to construct the cardiovascular system shown in Figure 1 below.

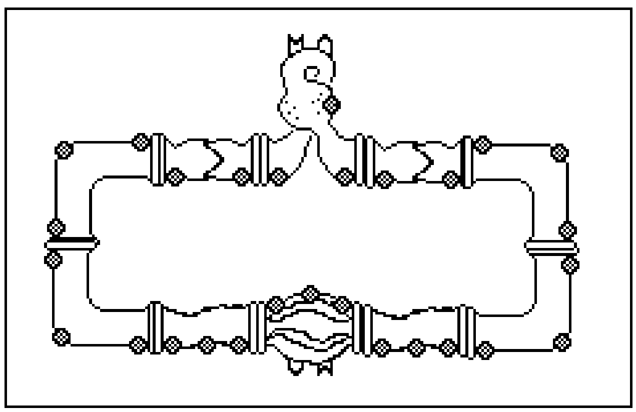


Figure 1: A simple cardiovascular loop with valves.

2. Run the simulation while carefully observing what happens on the screen. Run it as many times as necessary to answer the following questions:

- What is the direction of blood flow? (clockwise/counter-clockwise)
- When blood flows through a valve, is it open or closed?
- Do the two valves open and close at the same time, or do they open and close at different times? Why?

Exercise 2: Measuring Values

Sometimes it's hard to tell what's going on just by looking at the running simulation, since everything is happening so fast. This is especially true when you are trying to compare certain flows or pressures to eachother. This exercise focuses on the use of a gauges to measure and record blood flow or pressure at various places in the construction.

1. Modify the system you originally constructed by attaching gauges at the places marked in Figure 2. Attach a pressure gauge to the heart at point A, and flow gauges at points B and C.

2. Convert the gauges into graphs by double-clicking on each of them. Move the three graphs so that they are aligned vertically, one above the other. This will make them easier to compare.

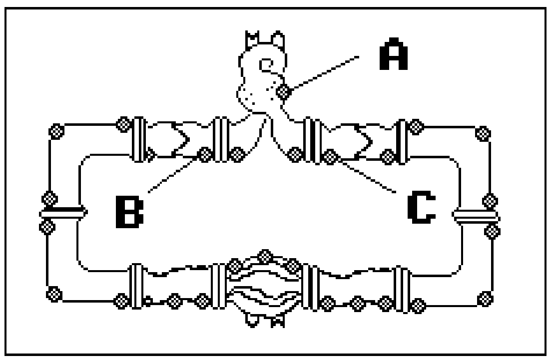


Figure 2: Where to attach the gauges

3. Run the simulation to answer the following questions:

• Check all of the following statements that are true:

_____There is never *any* flow past point B.
_____There are times when there is flow towards the heart at point B.
_____There are times when there is flow away from the heart at point B.
_____Though the amount and direction of the flow may vary, there is always some flow past point B.
_____There are times when there is no flow at point B.
_____Blood always flows *towards* the heart at point B, or not at all.
_____Blood always flows *away* from the heart at point B, or not at all.

• When does blood flow *towards* the heart take place at point B? Check all that apply.

_____ Never
_____ As pressure in the heart increases

______ As pressure in the heart decreases
______ When pressure in the heart is low and steady

- When does blood flow *towards* the heart take place at point C? Check all that apply.

_______ Never
_______ As pressure in the heart increases
_______ As pressure in the heart decreases
_______ When pressure in the heart is low and steady

APPENDIX D

COMPLETE DATA RECORD

Part of the difficulty of developing a novel comparative methodology like Breakdown Analysis is that, because their are no similar efforts to look to for guidance, it is not immediately obvious what characteristics of the interactions analysed might be relevant to the comparison of breakdown behaviors undertaken in this study. A number of characteristics of interaction related to the various categories of breakdown were collected and explored as ways of comparing the communicative efficacy of the three environments, but were not found to be useful and were never used. For example, the possibility of measuring the severity of breakdowns was explored; whether or not participants' were directing their gaze at the same space was noted for each instance of breakdown; the amount of time spent by participants on each part of the task was noted. For the most part, these additional data were not found to be relevant to the analysis presented in this work. For the benefit of interested readers who may want re-examine the data from different analytic perspectives, Table D.1 gives a summary of all the data culled from the transcripts.

Table D.1 is organized as a database in which each row describes a single subtask (see Chapter III) within a single session. Each session consisted of four subtasks with a total of four sessions analyzed for each of the three communication environments, yielding a total of 48 rows of data. The dark horizontal lines in the table mark the boundary between the data from interactions that took place in each to the three communication environments explored. The meaning of column headings is given in Table D.1.

Table D1: Definition of Terms used in Table D2.

Term	Meaning
Session and Task#	Specifies the session and subtask described by that row of data. Recall that the overall task (see Chapter III) was divided into four subtasks embodying distinct phases of the task-solution process.
Time	Gives the time taken by participants to complete the subtask.
RS-M and RS-R	The number of times each participant turned to gaze at the other participant during this task. Obviously, this metric applies only to the Copresent and Audio-Visual Scenarios since participants had no visual access to the other participant in the Audio-Only environment. For the audio-video scenario, the number of gazes at the remote video monitor was recorded.
Verbs-M, Verbs-R, uttr-M, uttr-R, C-trans, T-trans.	One avenue explored during (and ultimately abandoned) during the analysis was to attempt to “normalize” the amount of breakdown behavior observed in each scenario by somehow quantifying the notion of “opportunity for breakdown to occur” for each category of breakdown. The number of verbs (Verb-R,Verb-M) used by each participant was used for Reference breakdown, the number of utterances was used for Verbal turntaking breakdown, and the number of cursor control and topic transitions (C-Trans, T-trans) were used for Cursor turntaking and Topic breakdown, respectively.
Sev	Severity. The idea of trying to quantify the disruptive effects of breakdown was explored by counting the number of utterances in the verbal repair of a given breakdown.
POA	Point-of-Attention. For each breakdown, whether or not participants had a synchronous direction of gaze at the moment the breakdown occurred was recorded. Thus, the value in this column represents the number of breakdowns (out of the total observed for the task) which occurred as participants had *asynchronous* points of attention, e.g., one was gazing at the workspace while the other was gazing at the lab book.
Impl’d	Implied transitions. Each cursor and topic transitions was judged as to whether that transition was explictly marked in the verbal channel. This provided a way of assessing whether distributed participants compensate for limited communicative resources by increasing the amount of verbal control management.

RefBD, VerbBD, CursBD, TopicBD	Number of observed breakdowns. These columns denote, respectively, the number of Reference, Verbal turntaking, Cursor turntaking, and Topic breakdowns documented during that task.

Table D.2: Summary of all parametric data collected during the analysis of transcripts

Info			Parametric		Reference Management						Verbal turntaking						Cursor Turntaking					Topic Management				
Session	**Task#**	Time	RS-M	RS-R	Verbs-M	Verbs-R	Ttt-vbs	Sev	POA	RefBD	uttr-M	uttr-R	Ttl-Utt	Sev	POA	VerbBD	C-trans	Impl'd	Sev	POA	CursBD	T-trans	Impl'd	Sev	POA	TopicBD
AO2-FF	**1**	3.9	0	0	25	42	67	0	0	0	45	61	106	0	0	4	11	4	0	0	2	27	8	3	0	1
AO2-FF	**2**	2.62	0	0	20	37	57	3	0	1	27	35	62	0	0	1	2	1	2	0	1	18	3	2	0	1
AO2-FF	**3**	3.62	0	0	20	43	63	2	0	1	30	58	88	0	1	4	8	4	0	0	1	26	6	3	0	1
AO2-FF	**4**	4.63	0	0	20	81	101	12	0	2	41	81	122	1	3	7	1	0	0	0	0	21	1	5	1	2
AO3-FF	**1**	6.27	0	0	83	87	170	10	0	4	138	117	255	10	5	30	26	20	17	4	18	62	13	10	1	3
AO3-FF	**2**	3.3	0	0	59	49	108	12	0	2	57	57	114	2	3	10	14	14	0	0	0	18	4	10	1	2
AO3-FF	**3**	4.68	0	0	57	70	127	1	1	1	88	85	173	1	5	10	10	8	11	0	8	44	7	2	2	2
AO3-FF	**4**	7	0	0	102	105	207	0	0	0	119	109	228	3	4	25	14	14	0	0	1	37	5	13	2	4
AO4-FF	**1**	3.67	0	0	27	26	53	7	0	2	67	52	119	2	2	11	12	6	5	0	2	29	7	7	0	2
AO4-FF	**2**	5.02	0	0	33	61	94	15	2	4	81	98	179	2	3	10	2	1	0	0	0	37	9	10	1	1
AO4-FF	**3**	3.47	0	0	24	39	63	4	2	4	57	64	121	1	3	10	12	8	1	1	1	30	6	3	0	1
AO4-FF	**4**	10.6	0	0	65	149	214	3	1	1	174	215	389	5	7	32	8	7	0	0	1	55	11	1	0	1
AO5-FF	**1**	3.47	0	0	25	37	62	4	1	2	46	59	105	3	2	11	16	11	6	1	5	31	9	2	1	1
AO5-FF	**2**	3.1	0	0	18	25	43	2	1	1	30	34	64	0	1	4	3	2	0	1	1	9	5	0	0	0
AO5-FF	**3**	4.07	0	0	26	41	67	0	0	0	48	63	111	2	2	7	6	4	3	2	3	26	7	9	1	1
AO5-FF	**4**	6.9	0	0	63	99	162	2	0	1	92	114	206	1	2	16	5	5	0	0	0	26	9	15	1	2
AV2-MM	**1**	5.08	1	5	55	42	97	8	1	2	78	60	138	1	3	9	3	3	0	0	0	38	14	2	1	1
AV2-MM	**2**	2.93	0	2	55	29	84	15	2	5	49	43	92	2	1	7	4	4	0	0	0	19	3	0	0	0
AV2-MM	**3**	6.3	1	2	80	53	133	0	0	0	98	75	173	2	1	15	9	7	1	2	3	54	14	4	0	2
AV2-MM	**4**	3.98	1	4	65	44	109	0	0	0	86	62	148	3	1	9	5	5	0	0	0	27	4	0	0	1
AV3-MF	**1**	4.32	2	7	53	40	93	2	0	1	63	51	114	0	1	9	11	9	7	2	11	37	18	5	2	3
AV3-MF	**2**	7.12	8	12	111	70	181	10	0	4	107	80	187	5	11	23	11	11	3	0	5	37	6	19	3	5
AV3-MF	**3**	10.7	6	16	71	129	200	12	0	4	120	133	253	2	4	13	23	21	6	0	8	73	22	19	2	3
AV3-MF	**4**	7.5	19	16	97	69	166	10	0	3	94	92	186	5	8	22	10	10	0	0	0	29	3	10	0	1
AV4-FF	**1**	2.68	0	2	11	36	47	2	0	1	24	38	62	1	1	5	1	1	0	0	0	27	10	0	0	0
AV4-FF	**2**	1.18	0	0	11	15	26	2	0	1	17	15	32	0	0	1	0	0	0	0	0	6	2	0	0	0
AV4-FF	**3**	2.77	1	1	29	20	49	2	0	1	35	33	68	1	1	5	2	2	0	0	0	22	3	3	1	1
AV4-FF	**4**	2.27	2	3	31	19	50	0	0	0	27	23	50	0	0	2	0	0	0	0	0	15	1	0	0	0
AV5-FF	**1**	9.87	3	5	58	53	111	16	0	5	57	79	136	0	0	1	8	7	2	0	2	53	17	1	0	1
AV5-FF	**2**	1.83	0	2	7	8	15	0	0	0	8	9	17	0	0	0	0	0	0	0	0	6	3	0	0	0
AV5-FF	**3**	10.5	1	2	52	71	123	2	0	1	67	80	147	1	3	8	10	10	0	0	3	62	21	9	2	4
AV5-FF	**4**	5.12	4	6	16	22	38	0	0	0	22	34	56	0	3	4	1	1	0	0	0	16	7	5	2	2
FF2-FF	**1**	3.35	0	0	39	17	56	6	0	2	75	45	120	0	3	9	5	4	0	0	0	24	12	0	0	0
FF2-FF	**2**	2.32	4	1	16	12	28	0	0	0	38	29	67	0	1	3	2	2	0	0	0	9	2	0	0	0
FF2-FF	**3**	7.77	2	3	64	83	147	1	0	1	92	123	215	0	2	15	1	1	0	0	0	47	10	5	1	1
FF2-FF	**4**	8.75	1	5	131	91	222	0	0	0	156	123	279	3	1	21	10	8	2	0	2	50	11	5	0	2
FF3-MF	**1**	2.38	0	2	19	9	28	0	0	0	21	19	40	0	1	5	1	0	0	0	0	24	16	0	0	0
FF3-MF	**2**	4.25	6	8	49	25	74	1	0	1	44	28	72	1	1	4	0	0	0	0	0	21	8	0	0	0
FF3-MF	**3**	2.2	0	1	9	8	17	0	0	0	18	13	31	0	1	4	0	0	0	0	0	17	11	0	0	0
FF3-MF	**4**	3.05	0	3	26	19	45	0	0	0	33	36	69	1	1	5	0	0	0	0	0	17	5	4	0	2
FF4-MM	**1**	3.47	1	1	29	41	70	0	0	0	53	55	108	3	1	7	1	1	0	0	0	32	15	0	0	0
FF4-MM	**2**	2.4	0	3	28	32	60	1	0	1	35	37	72	4	3	13	0	0	0	0	0	12	1	0	0	0
FF4-MM	**3**	3.62	1	1	30	36	66	6	0	2	56	53	109	4	0	9	2	2	0	0	0	30	11	6	0	2
FF4-MM	**4**	2.63	0	2	41	34	75	2	0	1	50	40	90	1	2	10	0	0	0	0	0	19	5	0	0	0
FF5-MM	**1**	3.73	0	1	62	14	76	0	0	0	67	30	97	3	2	11	5	4	0	0	0	28	7	3	1	1
FF5-MM	**2**	1.97	1	0	37	11	48	0	0	0	29	13	42	1	0	2	2	2	0	0	0	12	4	4	1	1
FF5-MM	**3**	2.32	0	0	13	9	22	6	0	3	34	17	51	0	2	4	1	1	0	0	0	19	6	0	0	0
FF5-MM	**4**	5.17	0	0	94	17	111	2	1	1	94	37	131	2	3	5	1	1	0	0	0	28	1	1	0	1