

Design for conversation: lessons from Cognoter

DEBORAH G. TATAR,[†] GREGG FOSTER[‡] AND DANIEL G. BOBROW

System Sciences Laboratory, Xerox Palo Alto Research Center, 3333 Coyote Hill Road, Palo Alto, CA 94304, USA

(Received April 1990 and accepted in revised form August 1990)

When studying the use of Cognoter, a multi-user idea organizing tool, we noticed that users encountered unexpected communicative breakdowns. Many of these difficulties stemmed from an incorrect model of conversation implicit in the design of the software. Drawing on recent work in psychology and sociology, we were able to create a more realistic model of the situation our users faced and apply it to the system to understand the breakdowns. We discovered that users encountered difficulties coordinating their conversational actions. They also had difficulty determining that they were talking about the same objects and actions in the workspace. This work led to the redesign of the tool and to the identification of areas for further exploration.

Introduction

The Colab project was an ambitious attempt to provide computational support for group work, particularly for the support of small design teams working together in the same room. The project coordinated several technologies, including networked computers, video network facilities, and a specially designed room. Moreover, Colab had an elaborate and articulated model of the meeting processes it tried to support (Foster, 1986; Foster & Stefik, 1986; Stefik, Foster, Bobrow, Kahn, Lanning & Suchman, 1987a; Stefik, Bobrow, Foster, Lanning & Tatar, 1987b). This approach was exciting and important because it promised to permit significantly more effective ways of working. However, the observation of users working with Cognoter, the most developed Colab tool, showed that there were serious breakdowns in the system.

While trying to understand these breakdowns, it came to our attention that there was a potential conflict between the “interactive” model of communication proposed by recent work in psychology and sociology and the “parcel-post” model implicit in Cognoter. However, the importance of the difference between the two was initially unclear because the interactive model was created to describe situations (two-person, purely verbal communication, often without visual contact) that differed in important respects from the situation we had created in Cognoter. We drew upon sociological work studying the use of traditional representational media such as whiteboards to extrapolate from the interactive model and argue about its significance for our system. This analysis guided the redesign of the system.

This paper starts with an overview of Colab and Cognoter. We then describe

[†] Author's current address: Stanford University, Department of Psychology, Bldg. 420, Stanford, CA 94305-2130, USA.

[‡] Author's current address: Apple Computer Inc., 20525 Mariani Ave., Cupertino, CA 95014, USA.

some of the difficulties that our users had working with the system. Next, we present elements of the interactive and parcel-post models of conversation. To extend appropriate expectations to the Cognoter situation, we discuss the use of traditional representational media in meetings. This allows us to create a picture of the problems our users faced. Lastly, we talk about the implications of this line of thinking for the redesign of the system, for understanding more about communication, and as embodying techniques which are important for CSCW systems in general.

Colab and Cognoter

The Colab room was designed to enable the use of computers in meetings of two to five people. The room consisted of three specially designed tables arranged in a U-shape facing a large screen at the front of the room (Figure 1). Each table had on it a display, keyboard and mouse. Each display was connected to a separate processor. The processors were connected to one another by an Ethernet network. Additionally, the displays were connected to one another and to the large screen ("Liveboard") by a video network. The video network could be used to project any of the small displays on the Liveboard, as well as allowing any user to project another user's screen on her own display.[†] When the video network was being used to look at another station's display, the user lost access to the input devices and computational facilities of her own workstation.

Cognoter was software designed for the Colab to aid small work groups in the creation of a plan or outline. Cognoter "implemented" a three-part process of brainstorming, organizing the brainstormed ideas into sequences and groups, and evaluating them. As originally conceived, Cognoter was supposed to be a fairly

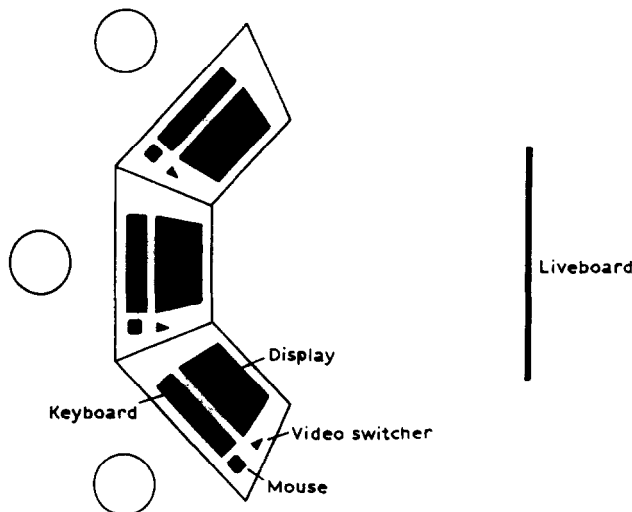


FIGURE 1. A diagram of the Colab room. An Ethernet network connected the computers, and a video network connected the displays and the large display ("Liveboard") at the front of the room.

[†] This video facility was intended primarily for use with software outside the Colab project.

direct translation of a process that we ourselves used in meetings with the whiteboard and other static representational media.

However, Cognoter took advantage of the Colab setting in several ways. To increase the effectiveness and efficiency of the meeting, the software allowed participants to work on their computers in parallel. They could each use their keyboard and mouse at any time to make contributions. We hoped that this would reduce the “production blocking” (Diehl & Stroebe, 1987) that is an impediment to group idea generation.

The fundamental unit in the Cognoter user interface was the item. Each Cognoter item consisted of an icon made up of a short catchphrase, usually limited to about 20 characters. Additional text could be associated with the item to explain the catchphrase. We thought of this additional text as an annotation to the item icon, although it could also be thought of as the content of an item. A special kind of item, a *group* item, could have other nested items associated with it.

Cognoter had two kinds of windows: *item organization windows*, whiteboard-like spaces used to display and move the icons, and *edit windows*, used to create and change both the catchphrases displayed in the item icons and the annotations associated with the items.

A major innovation of the project over other computerized brainstorming tools was to use the item organization windows to give users *shared workspaces*. The item organization windows appeared on every user's display at the same time. Each user's copy of an item organization window contained the same information and behaved the same way as every other user's copy. This approach and the many possibilities it raises are discussed in Stefik *et al.* (1987a).

To create an item, a user opened an edit window. When she completed the entry of the item, the system created an icon to be placed in an item organization window for everyone to see (Figure 2). Once created, the item was equally available to all users. By clicking on the icon, a user could either drag it to a different location or open an edit window and change it. A user could make a group item expand so that its associated group of sub-items appeared in a new item organization window. Icons could be moved within and across item organization windows. Items, nested groups and annotations allowed users to create what may be thought of as an annotated graph of ideas, with items and annotations at the leaves and groups at the nodes (Figure 3). This graph could be useful in its own right as a way of representing a complex problem space, or it could be used to generate a linear outline.

In theory, the computational representational medium seemed to combine the features of several tools in current use (Tang, 1989; 1990). As on a whiteboard, each person shared the same up-down orientation or perspective on the material. As on large sheets of paper (as used by architects and other designers), each person shared a close physical relationship to the representational medium. Lastly, as on a computer, information could be handled flexibly; even if the amount of material was quite large, editing, saving, restoring and printing were all possible and easy.

We believed that Cognoter changed people's patterns of work and even what happened in each person's head: “Cognoter... divides the thinking process into smaller and different kinds of steps that are incremental and efficient” (Stefik *et al.*, 1987a, p. 35). We also expected it to change the way people relate to one another in meetings. However, initially we considered design trade-offs with regard to meeting

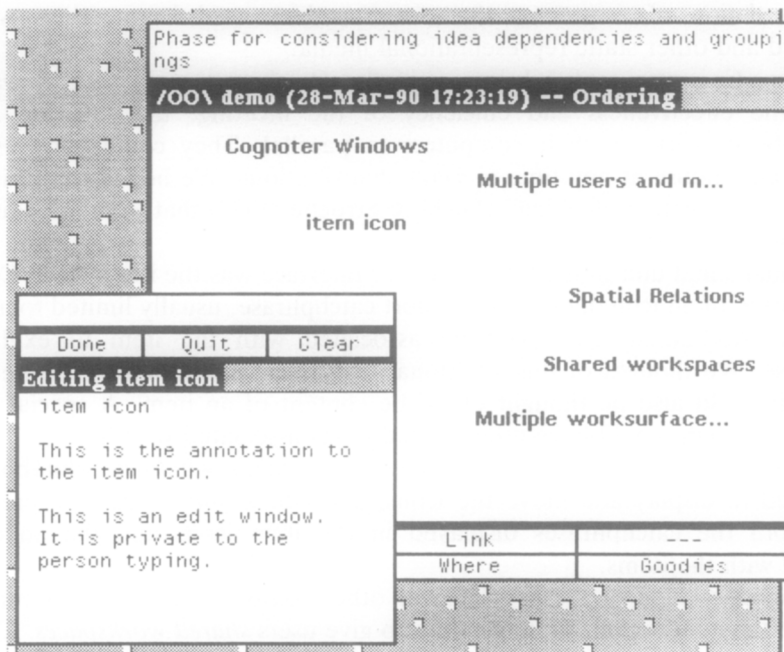


FIGURE 2. Item organization and edit windows. Item organization windows were shared and allowed users to arrange item icons. Edit windows were used to create and modify items.

processes only as they seemed to fix problems such as overlooked ideas and inefficiency. We did not consider whether the interventions we proposed manipulated resources at the level of basic human communication.

Experiences working with Cognoter

We had substantial experience using Cognoter in a series of working sessions with one or more members of the Colab group participating. Reports from these experiences were mixed, with a number of positive responses, but many unhappy comments. However, meetings can be good or bad for many reasons not related to the technology. Although preliminary observations of Cognoter use (Stefik *et al.*, 1987a) anticipated the problems detailed here, these observations were hampered by the impossibility of seeing the details of work between three or four users working on separate machines. The observations were also, as it turned out, hampered by the ability of people who were very familiar with the performance characteristics of the software to compensate for its problems.

To gain an understanding of what happened to "real" people, we asked two outside groups to work with Cognoter. Each group consisted of three long-term collaborators who were familiar with the editor, window system, and mouse conventions Cognoter used. Each group was asked to brainstorm about a subject of their own choosing that would be useful for their own work. Both groups worked for two two-hour sessions. To solve the observational problem, we videotaped them at work and kept a record of all messages sent between the different machines in the session.

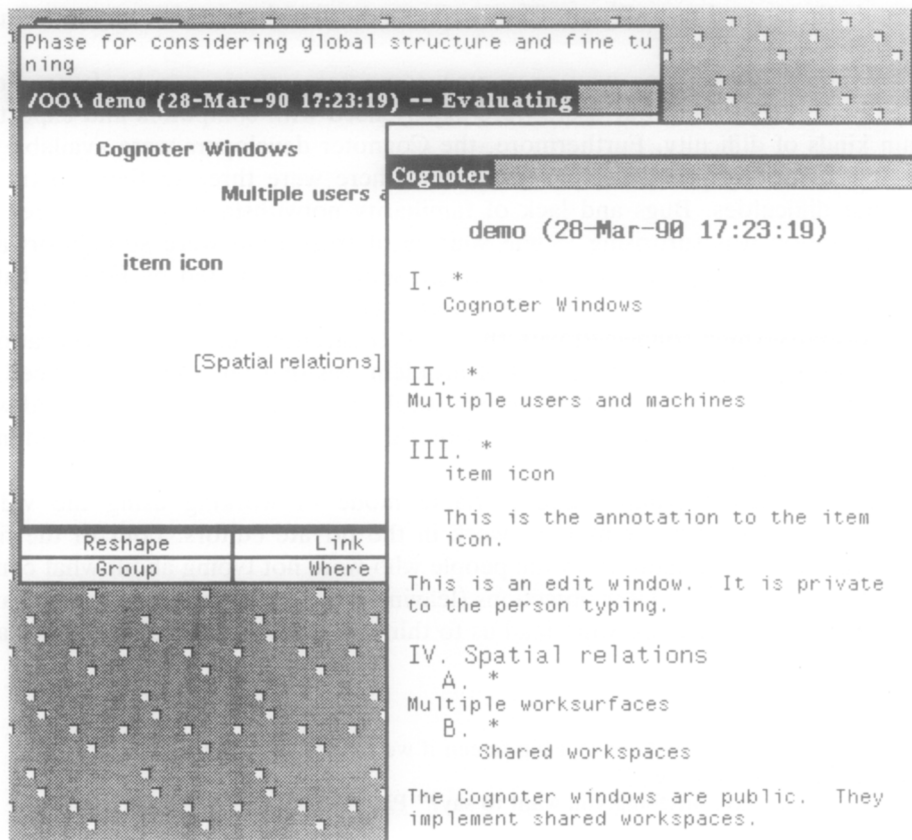


FIGURE 3. Items, annotations, and groups created an annotated graph. The graph could be used to generate a linear outline.

Our trial groups encountered serious problems. In one group, work proceeded in two phases. In the first, each person started an edit window and worked in it. They hardly talked at all and did not look at one another's work. In the second phase, they stopped using the system altogether and resorted to working together with a pad of paper. The other group managed to find a successful way of using the tool by using the video network to look at the screen of whoever was typing, thus employing the shared video workspace instead of the shared computational workspace. This solution worked rather well for them, and they ended up pleased with some aspects of the tool. However, this meant that they lost one of the chief features of the system, the ability to switch typists easily. Far from attaining the expert, fluid trade-off, seen in designers working around a large sheet of paper, they had to spend quite a bit of effort negotiating who would type next.

Both groups bypassed the computational shared workspace either by working privately and then doing the group work on paper, or by giving up on their input devices and using the video connection to create a visually shared workspace. When they tried to use the shared workspace created by the software, they found it so frustrating that they put their heads in their hands, raised their voices, and

ultimately threatened to walk out. They expressed astonishment that anyone would build such a system.

There are many reasons why prototype computer systems can be frustrating, especially to novices, but our users were experienced with computers and expected certain kinds of difficulty. Furthermore, the Cognoter developers were available to help them with any problems that arose, and there were three of them to try to figure out difficulties. Bugs and lack of familiarity notwithstanding, the degree of their frustration was surprising. Some sources of frustration were straightforward once observed and led immediately to design solutions, reported by Foster, Tatar and Bobrow (Unpublished data). However, there were two major classes of problems that seemed connected with the worst frustrations and whose implications required more thought. The first kind of problem was that our users felt a need to see things in the workspace that the system would not let them see. The second was that they mistook references in one another's speech or actions and could not resolve the difficulty satisfactorily.

Before the second group found a viable mode of working using the video network, they made five attempts to work in the private editors. Four of the five attempts evoked a complaint from the people who were not typing about what could and could not be "seen." User objections (leaving out the tone of voice, gesture and surrounding detail which are what lead us to think of these as particularly important reactions) included:

- "Why can't I see that?"
- "I don't see what use it is to have a big screen if we can't all contribute to it."
- "Click DONE so I can see it."
- P1: "P2, do you have anything you want to say?" P2: "I won't be able to see it up there, right?"†

These objections are united by participant confusion and difficulty in seeing what they needed to see. Even the second comment, which looks at first glance like a complaint about the distribution of information between the large screen and the individual displays, reflects confusion about what could be seen. (Since the large screen at the front projected one of the individual displays, they could contribute to it simply by contributing to the shared workspace. The fundamental objection concerned what could be seen in the shared workspace and it is particularly telling that they did not immediately recognize that a shared workspace was projected.)

The fifth case, although relatively unproblematic, is still instructive. In this case, one user started to take notes on what the other two were talking about. Subsequently, he read his notes back to them, adding "Well, you can all see this". They could indeed all see it by looking at the large screen at the front of the room which had been set so that it displayed his screen. Yet, even in this unproblematic usage, the user felt he had to read aloud what could have been read by each participant separately, and to comment on the fact that they could have seen it. In other words, he had to take action to bring it back into the conversation.

These objections establish the problem that users could not see things. However, the objections also raise questions. What exactly do the users expect to see? In a meeting involving a whiteboard, one participant might be looking out of the window,

† "P" followed by a number is used to denote participants.

or be taking notes, or whispering to someone else, and not see a gesture or drawing as it is put up. Yet this lack of attention does not usually cause extreme frustration or a breakdown of the situation. What was significantly different in the Colab situation?

We also saw a second kind of problem that required explanation. This was a problem with mistaken reference. As Bly (1988) has documented, deictic references, such as pointing to the screen and saying “there”, “this”, or “that”, persist even when such behavior has limited utility, as when we are on the phone. However, any ambiguity raised by this practice is by and large handled without remark. In fact, deictic reference is often ambiguous even in face-to-face conversation. Nonetheless, the ambiguity our users experienced appeared to be more extreme than in these other cases. For example, our users had a problematic breakdown in which they appeared to have quite different interpretations of the word “that”. A summary account of what happened follows:

(1) P2 was using the video network to look at P1’s machine. There was a general discussion about creating a new item;

(2) P1 hit a mouse button, thereby creating an edit window for entering a new item;

(3) P3 suggested that P2 should type instead of P1;

(4) P2 went to switch her display back to her own machine;

(5) P1 looked at her screen, appeared surprised, typed four characters (the title of the item) and moused the DONE button, thus sending it to the others.

(6) P2 found that her display had gone black (into idle) and was confused, “What! Who did that?” At virtually the same time, the item that P1 had created appeared on P2’s and P3’s screens.

(7) P3 “P1”;

(8) P2 (hitting a space bar which caused the machine to come out of idle) “Oh, it was my fault”.

(9) P1 “I did not!”

(10) P3 “P1, Let P2 type!”

(11) P1 “I am!”

(12) P2 “I’ve forgotten what I was going to type”.

P2 said “Who did that?” referring to the fact that her machine had gone into idle. P3 interpreted “that” to mean the thing that had changed in his environment, which was the appearance of a new item. He had not caused this new item to appear, so he replied that P1 had caused it. P3 ended up with the wrong picture of P2’s complaint and therefore P1 was unjustly accused of not permitting P2 to type.

In fact, the participants/collaborators moved on without sorting out what had happened, and it was only through careful reconstruction that we as analysts came to understand the sequence. Since both our trial groups consisted of long-term collaborators, the amount of interpersonal damage was probably small, but it did represent a substantial disruption to their work. Furthermore, the incident, combined with the fact that both groups declined to use the computational shared workspace, was telling. It suggested that in this situation people could not rely on familiar mechanisms for coordinating or managing ambiguity. The technology was not increasing their efficiency in this respect.

To proceed, we needed to understand what was causing the difficulties our users experienced. We needed to understand what they needed to see and what factors contributed to the increase in ambiguity. However, since this was not intended as our major line of research, but rather as enabling work for other more central interests, we faced the challenge of trying to work as much as we could with materials already in the literature (rather than constructing a research project to explore this one set of issues). The challenge that we faced is a general one because CSCW systems almost always touch on many different research areas, not all of which can be pursued actively. Designers need approaches that maximize their use of appropriate existing research.

Models of conversation

The fields of conversation analysis (see Goodwin & Heritage, 1990 for an overview of the field) and psycholinguistics (see Clark & Wilkes-Gibbs, 1986; Clark & Schaefer, 1987, 1989; Schober & Clark 1989; for the most relevant aspects) offer a model of conversation which seems highly pertinent to the difficulties our users faced. They present what we may term an "interactive" model of communication. The interactive model emphasizes the notion that conversation is a highly coordinated activity in which meaning is attained and affirmed using a number of mechanisms that have context dependent functions. By contrast, Cognoter implemented and thus supported what we may term a "parcel-post" model in which communication is delivered in parcel-like units.

INTERACTIVE MODEL

The starting point of the interactive model of conversation is the observation that conversation does not consist of one person making a complete utterance while the other person waits passively. Both participants are active even when only one is actually speaking. Thus people nod, complete or reshape one another's phrases, and say "uh-huh" (Duncan, 1973).

A second major point of the interactive model is that the function of an utterance is context dependent. Each conversational move involves not only its own contents but a projection of what the next move will be. For example, one of the resources that people have available in conversation is the *noun phrase*. There are different types of noun phrases (Clark & Wilkes-Gibbs, 1986), the most common being the *elementary* noun phrase, as in "the green cup on the bookshelf" uttered as one tonal sequence. When a person utters "the green cup on the bookshelf", she is projecting that the phrase will be accepted without comment. Other types of noun phrases, e.g. "the whatchamacallit", set up other expectations. If she alters the phrase by the use of a *try-marker* (Sacks & Schegloff, 1979), a tonal, non-lexical request for judgement about its acceptability,[†] she is indicating that the preferred next move is a listener response indicating whether his meaning is clear. In the case of an unmodified elementary noun phrase, a non-response complies with the projection and asserts that the phrase has been understood. If the phrase is modified with a try-marker, a non-response indicates trouble, perhaps that the other person is no

[†] A try-marker sounds similar to a question, but it is not a request for agreement on contents, just agreement that the phrase is understood.

longer listening. Furthermore, a response after a pause is not the same as an immediate response. For example, it may set up a question about whether the listener is attending or beginning to signal dissention (Pomerantz, 1984; Pomerantz, 1978; Goodwin, 1980).

Listeners have the ability to make statements explicitly or implicitly which add to, accept, reject, question or modify what the speaker has just said. In the interactive model, this listener response is a crucial part of conversation. In fact, Clark and Schaefer (1987, 1989) have gone so far as to advance the notion that the basic unit of conversation consists of two parts, a *presentation* and an *acceptance* phase. Together the two constitute a *contribution*. Since, as mentioned above, non-response is in fact a statement, the ability to perform the acceptance portion actively is crucial.

The interactive model also draws our attention to the importance of mid-course corrections (Goodwin, 1981; Schegloff, 1981; Clark & Wilkes-Gibbs, 1986). Mid-course corrections happen when speakers or listeners adjust in midstream, either of their own accord or in response to something in the environment. For example, while uttering the phrase "the green cup on the bookshelf", the speaker may note the listener's confused expression and attempt to clarify by adding "in the library". We know that people can succeed in conversing without some elements of this because they succeed in conversing over half-duplex phone lines and with computer TALK systems which may be serial. However, most people find this quite unpleasant and, in face-to-face conversation, mid-course corrections, whether by the speaker or the listener are endemic. According to Clark and Wilkes-Gibbs, mid-course corrections are yet another way that conversationalists obey the general goal of "reducing the collaborative work" of holding the conversation.

In the interactive model, conversation is structured around the work that both participants must do to establish that they are talking about the same things. One aspect of this is establishing what objects they are referring to. To study this problem, Clark and Wilkes-Gibbs (1986) asked pairs of subjects to sort a deck of cards with non-representational pictures on them. The subjects each had their own deck and could not see one another or the other person's deck. The *matcher* was supposed to put the deck in the same order as the *director*. On average, the first references to each card took close to four turns at speech by the director. However, people's ability to refer verbally to particular objects developed through shared experience. As they repeated the task, the references they used to the cards became more compact. They were simplified and/or narrowed to the crucial components. Thus, the second trial took on average two turns and the third averaged at close to one. Furthermore, the number of words per turn at talk declined over time; "the one that looks like a skater with one leg kicked back" turned into "the skater" (Clark & Wilkes-Gibbs, 1986).

Furthermore, in general people preferred to refer to the task items with permanent descriptions rather than temporary views. A permanent reference is one such as describing a drawing as "the rabbit". This contrasts with temporary sorts of references such as "the one we got confused on last time". A reason for this preference may be that permanent descriptions permit successive refinement of the reference.

In the interactive view, conversation is a complex, highly coordinated process in

which conversants seek mutual understanding through the coordinated presentation and acceptance of a variety of lexical and non-lexical statements. Projections of the next conversational move and active listener response are crucial for this activity. The abilities to make mid-course corrections and abbreviated references are highly desirable. Time plays a crucial role.

PARCEL-POST MODEL

Although Cognoter permitted people to talk with one another, and thus employ their normal resources, its textual component may be described as embodying a parcel-post model of communication. The qualities of a parcel-post model are that items are packaged and sent by the speaker, and then unpackaged and decoded by the receiver. An additional component of the model is that if the receiver does not open his "mail" right away, he may end up with a bunch of stuff with no particular order.

The differences between these models suggest that one class of problems may occur in coordinating such interactive features as mid-course corrections, projections of the next move, and listener response. A second class of problems may occur in determining that both people are talking about the same thing.

Problems applying the interactive model to Cognoter

The differences between the interactive and parcel-post models suggest that our users had difficulty because they could not accomplish necessary or highly desirable activities described by the interactive model. However, we cannot immediately conclude that this explains our users' problems, because the interactive model arose from studying situations that differed in two important respects from the situation we created in Cognoter: the number of people involved in an interaction, and the type of communication involved. To understand whether and how we would expect the interactive model to generalize to Cognoter, we must understand the impact of these differences.

The first issue arises because most of the work that led to the interactive model studied conversation between only two people. Our Cognoter experiences involved three people. In recent years, there have been some studies of multi-party interaction which extend but do not refute the two-person model (Goodwin & Goodwin, 1990). While these studies provide some assurance of continuity between the situations, they have not concentrated on identifying the differences between the two party and multi-party conversation. They leave open the possibility that, for example, the expectation of response is lessened in the multi-party case. Nonetheless, conversation with three people is, by and large, unlikely to simplify the possibilities or render resources for achieving shared reference less necessary. Although there are unanswered questions, we may safely assume that any factors that seem likely to cause difficulties in the two-person situation will be only more likely to do so in the three.

The second problem originates in the recognition that many forms of communication exist which are not conversational. For example, the parcel-post model works perfectly well for letters. Since people are creating text in Cognoter, we initially assumed that the parcel-post model would suffice. Why should we now attribute their difficulties to conversational impediments?

Clark and Wilkes-Gibbs (1986) propose a highly relevant distinction between conversation and literary communication. They propose that the principle of *mutual responsibility* holds in conversation:

The participants in a conversation try to establish, roughly by the initiation of each new contribution, the mutual belief that the listeners have understood what the speaker meant in the last utterance to a criterion sufficient for current purposes (Clark & Wilkes-Gibbs, 1986, p. 33).

They contrast this with the principle of *distant responsibility*:

The speaker or writer tries to make sure, roughly by the initiation of each new contribution, that the addressees should have been able to understand his meaning in the last utterance to a criterion sufficient for current purposes (Clark & Wilkes-Gibbs, 1986, p. 35).

The principle of distant responsibility applies in the many situations—writing a paper, sending e-mail, giving a lecture, broadcasting on TV, dictating a tape to be sent in the mail—that are distinguished from conversation less by the medium carrying the communication (paper, airwaves, Ethernet) as by a relative lack of co-production and therefore of time constraints in the preparation of the communication. In other words, conversation is distinguished from literary forms of communication by the amount of work to ensure understanding that is done *within the time frame* of the actual communication. Cognoter differs from literary communication because Cognoter is not the sole or primary carrier of the communication. Cognoter differs from conversation as we have described it because it involves writing activity. In fact, Cognoter is neither of these, but rather a medium for representation in relationship to the conversation.

A large body of work has been done exploring the relationship between conversation and writing activity in traditional representational media such as whiteboards and large sheets of paper. This work provides evidence that similar constraints apply when working with representational media as with unadorned conversation, and that similar resources are necessary or desirable.

Traditional representational media

Suchman (1988) and Tang (1989, 1990) and Bly (1988) have studied traditional representational media such as whiteboards and large sheets of paper. We can draw on their work to extrapolate from the interactive model to understand how we might expect the parcel-post model of Cognoter to cause users trouble.

Suchman and Tang give evidence that, since writing and drawing activity interacts with conversation, coordination is as important when using these media as it is in conversation without them. Suchman (1988) describes the way in which turns at "... the board may be used in taking and holding the floor, or in maintaining some writing activity while passing up a turn as talk. Writing done during another's talk may (a) document the talk and thereby display the writer's understanding, (b) continue the writer's previous turn or (c) project the writer's next turn, providing an object to be introduced in subsequent talk". The writing actions are neither appendages to the verbal conversation nor independent of it. Writing and talking is intricately bound together in a similar fashion to the way a statement and its response are bound

together. However, the coordination issues are even more complex because a person can write while talk is occurring. There are therefore more kinds of moves to be made.

The fact that textual items persist makes it tempting to believe that time is less relevant in shared writing and drawing activities than in purely verbal conversation. "Recording information" is usually accounted the chief reason to use a whiteboard and this activity is considered to be independent of the conversation. However, in practice, only a small percentage of the activity in relationship to a whiteboard is recording information (Tang 1989). It is important therefore to look at other functions of work in the shared medium.

Suchman (1988) distinguishes between those actions in reference to the whiteboard that involve the production of textual or graphical objects and those that involve the use. Our construction of this distinction is that whiteboard items hold a dual status as elements *in* the conversation and elements that may be conversed *about*. During production (when they are being created) they are typically, but not always, elements in the conversation and, like verbal conversation, understood according to precise context. Thus, the moment a person starts to write helps determine whether the writing action acknowledges what has already been said (or written) or ignores it, and whether the writing represents agreement, clarification, informed disagreements, a side activity, or an attempt to bring up a new topic. Once created, the textual items are similar to other physical objects such as tables and chairs in the environment and may be conversed about.[†] As objects in the world, they are not subject to timing considerations. However, participants may have to do work to ensure that they are talking about the same objects in the world, and time is a factor in that work.[‡]

Recording information for future use is certainly an important function for representational media. However, even while recording information, participants are not free from timing considerations. Tang (1989) reports that the length of time recording information takes, presents a "challenge" for the group. This challenge can be handled in a number of different ways. They may wait for the recorder to finish, occupy the pause with individual work, or move on to another topic. Any of these activities may be accompanied by talk, preceded by it, followed by it, or free of it. Drawing upon the Clark and Wilkes-Gibbs distinction between conversation and literary communication, we may speculate that the challenge arises because the person doing the recording has in an important sense stepped out of the conversation. She is engaged in an essentially literary endeavor in which she must take the time to anticipate what the group or someone else will need to know later on. Meanwhile, the other participants may compensate, e.g. by monitoring the activity (but not always).

A further issue arises when we consider whiteboard items as elements to be referred to, as objects in the world. In everyday conversation and in whiteboard use,

[†] The situation is a bit more complex than this. Since production takes time and items once created may be modified, which is a kind of production, a given item on the whiteboard has the potential to be both an element in the conversation and to be conversed about as an object. People intermix actions in relationship to these different aspects of whiteboard elements smoothly.

[‡] Furthermore, we expect that work to be increased over the work people have to do to talk about tables and chairs, because the objects which are in general not familiar, represent other complex ideas, and may not be as easily recognized.

deixis, the ability to point or say “this” or “that” with reference to the objects in the environment, is omnipresent. The success of a deictic reference depends on shared knowledge about the position of an object. This suggests that one reason that whiteboards are so useful is they give a highly salient quality or position to an idea. The salience of position is demonstrated by instances of people referring to the spot on the whiteboard where a particular idea, now erased, once was represented (Suchman, 1988). Clark and Wilkes-Gibbs do not discuss deixis. However, they do note that people prefer permanent qualities to temporary ones. In their study, location was considered a temporary feature. This makes sense because they were talking about stacks of cards that were being continually moved in the course of the task. However, the prevalence of gesture in relation to whiteboard objects and the salience of position, provides evidence that the location of whiteboard objects is regarded as a permanent quality for the purposes of the conversation. Thus, threats to positional information will make object identification more difficult.

The study of traditional representational media contains ample evidence that people use the basic “interactive” conversational paradigm. It is highly structured and dependent on both time and context. While there may be interesting modifications in response to different representational media, we have every reason to believe that the same resources are available for projecting the next move, for making and obtaining listener response, and for mid-course correction. The whiteboard provides increased facilities for determining what objects are being discussed, and the success of this facility appears to be dependent on positional information, as well as on the contents of whatever is written.

Cognoter problems

Cognoter differs from whiteboards and large sheets of paper in important respects: items may be easily rearranged; people work with keyboards; and the participants have not only displays instead of boards, but separate displays. Still, Cognoter is like a whiteboard or large sheet of paper in several important respects: crucially, it allows the visual presentation and inspection of all items by all participants; it allows items to be created and pointed at; and it allows items, though of a limited nature compared with whiteboard elements, to be arranged freely within the available space.

We have drawn a picture of conversation in relationship to a representational medium that emphasizes interactivity and coordination. This picture also contains the notion that an important utility of the representational medium lies in the ability to refer to objects in a succinct way. We have evidence from Cognoter that users could not “see” what they wanted to see, and some evidence about mistaken reference.

To have enough confidence that the differences between the interactive model and the parcel-post model substantially account for the difficulties our users faced, we want to have a picture of how we might expect the interactive processes that the users were engaged in to play out against the system with which they were working. We cannot have a complete model because we don’t know all the resources that people have available or their significance at all moments. Nonetheless, demonstrating a severe blockage of the process components that we have identified, both

argues for the significance of these processes and highlights specific implementation decisions in the system.

In light of this, eight design decisions which we must consider as possible contributors to the difficulties our users had are as follows:

- **Separate screens:** We gave each user a separate screen. They were not in fact looking at the same place. This meant that gaze and gesture information was reduced.
- **Lack of sequentiality:** There was no marked position where the next icon would appear or any way of determining the order of contributions.
- **Short labels:** Icons could only be short phrases. This meant only a small amount of information per item could be viewed by the group together.
- **Anonymity:** All changes were anonymous. The results were delivered to others with no indication of who had made them.
- **Private editing:** Editing of item text was accomplished in private editor windows. The results were delivered to the others wholesale. The catchphrase (and therefore the icon) could be changed entirely. Alternatively, the item could retain the same catchphrase but the annotations which gave it a particular meaning could have changed.
- **Unpredictable delay:** Changes showed up on other people's screens after an unpredictable delay. Sometimes they showed up as quickly as a third of a second but at times it took as long as 20 seconds.
- **Private moving:** Moving icons was accomplished privately. On other people's screens, an icon would disappear from one position and appear in another. The object could lose its identifiable position.
- **Tailorable windows:** Users could tailor their screen individually by moving windows around and by changing their size and shape. Item organization windows could appear in different places in different sizes on each person's screen.

With these design issues in mind, we revisit the Cognoter situation and attempt to describe what we believe the users faced and why these factors contribute to a problem for them. We divide this discussion into two sections: coordination problems and reference problems.

For the purposes of this discussion, the person making the contribution is referred to as "he" and the person responding as "she". The "speaker" is used for the person making the contribution, while "listener", "recipient", and "responder" are used for the person responding.

COORDINATION

Cognoter users are presented with a choice of media; they can choose to communicate: verbally, through text, or by using a combination of the two.[†] Like all participants in a conversation, they collectively face the need to (1) produce contributions,[‡] (2) recognize contributions as such, and (3) make responses. Although by definition making a contribution must be a positive action, response may be made either by positive action or by a non-response. Verbal and writing activities need to be coordinated with one another if made by the same person, and

[†] For the purposes of this discussion, we treat these as distinct choices. However, people presumably intermix them smoothly.

[‡] Although Clark and Schaefer (1987, 1989) have used the term "contribution" to refer to the whole process of saying something and getting a response, we are using the term contribution more colloquially to refer to the spoken or written utterance. They use the term "presentation" for the utterance. "Presentation" is unsatisfactory for our purposes because it sounds as though there is a slice of time in which the listener is purely an "audience".

responses need to be understood in relationship to the contributions that evoked them.

Producing contributions

A person making a contribution in Cognoter has a choice of media. Making a spoken contribution in the Cognoter situation is hardly different from making a spoken contribution in the other situations we have discussed.[†] However, if the user attempts to include Cognoter in his action, he faces certain difficulties. A purely textual contribution, since it is made privately (private editing), does not in itself contain the elements of a bid for the floor, in the way starting to write at a whiteboard sometimes does.

The speaker may attempt to remedy this by accompanying his writing activity with speech. However, private editing and unpredictable delay mean that the textual and verbal elements of the contribution will be extremely hard to coordinate. On a whiteboard, users can perform different kinds of actions. We've observed that one action, namely recording information, presents a challenge for the other participants, which they handle with a variety of strategies. In Cognoter, even acts which do not involve recording information present a similar challenge insofar as they are invisible. It is as if the person, rather than making a bid for the floor, had simply dropped out of the conversation in the same way they have to when recording information in traditional media.

While the challenge of coordinating talking and writing is similar to the challenge the group faces when handling recording information on traditional media, their resources are not in all cases the same. If the speaker speaks as he starts to type and if his speech succeeds in capturing the attention of the group (i.e. if it is timed carefully enough despite the distraction of typing), then the group faces options similar to the traditional case (although the listeners may be waiting in situations which would not normally require them to wait: recall that our users asked unhappily "Why can't I see it?"). However, if the speaker waits until he has finished typing to speak, he risks losing any projectible connection between the time he initiated the typing and the time the message is received. Furthermore, because of separate screens, unpredictable delay, and lack of sequentiality, there will never be a particular moment at which the speaker knows that his text item has been received by everyone and is being looked at.[‡] Lastly, even if the recipients see the item come up, if more than one person was typing, they must figure out who's commentary matched the new (anonymous) item.

This lack of coordination means that the speaker is hampered in several ways. First, he cannot make or obtain mid-course corrections. Even if the speaker does adjust his writing activity in response to something the listener has said or done, this adjustment loses any meaning as acknowledgement or anticipation of listener reaction. The inability to evoke mid-course corrections increases the speaker's burden to complete the work of the phrase by himself, and increases the total work of the group since his extra work could in many cases be avoided.

Second, even once the contribution is complete, he encounters difficulties

[†] Contributions are not, in general, really made against a clean slate.

[‡] As with response, overt work can be done to make everyone look at the item, but this is a far cry from increasing the efficiency of the conversation or the meeting.

projecting a preferred response. Two resources we have identified as important in conversation are try-markers and pauses. These are strictly verbal. The absence of coordination between talk and writing means that try-markers and pauses are ineffective in relation to the computational medium.

Mid-course corrections, try-markers and pauses certainly do not represent the complete set of resources that speakers have available in conversation. However, their absence is quite a significant loss.

Recognizing contributions

We expect from the analysis above that the speaker will have to do more work to make a contribution in relation to Cognoter, than with traditional media or in simple conversation. However, the situation is even worse for the person who attempts to use the system to respond to another's contribution. First, she must know what to respond to and that itself presents difficulties.

The ability to know what to react to and when is reduced. If the speaker is typing and talking, the listener can respond at the appropriate time to either the verbal or the textual component of the turn. Since these are not coordinated, she must choose between them. If she has enough information to respond to the verbal component, then she has the difficulty of timing her response to avoid competing with the typing. If she waits to receive and assess the text, she has the difficulty of spotting it on the screen (lack of sequentiality). If she does spot the item, she must determine that no one else has priority making a response (lack of sequentiality, private editing, unpredictable delay). Furthermore, since people may well occupy their time making their own contributions, even when one contribution is detected, she may have trouble deciding on an order for response.

Additionally, since it is extremely difficult for the speaker to establish a connection between verbal and textual matter, the listener is likely to see the item in the absence of mid-course corrections, try-markers and pauses. This means that the phrase will be seen as "elementary", that is projecting assurance that it can be understood. She will therefore work hard to make sense out of it. This increase in her work is significant because even in everyday conversation, "the heavier burden usually falls on the listener, since she is in the best position to assess her own comprehension" (Clark & Wilkes-Gibbs, 1986, p. 34). Furthermore:

When the speaker utters I just found the keys, marking the noun phrase as elementary. . . the listener is under strong pressure to accept it. After all, the speaker marked it as elementary; so he must believe it to be adequate for current purposes. If she rejects it, she risks offending him by indicating that it wasn't adequate. She also risks revealing her own incompetence if indeed it should have been adequate. Finally, like the speaker, the listener wants to minimize collaborative effort—to avoid extra steps in the acceptance process—and that too puts pressure on her to accept. All this encourages her to tolerate a certain lack of understanding, even to feign understanding when it is not justified. She may do this trusting that the holes will be filled in later, or that they won't have serious consequences (Clark & Wilkes-Gibbs, 1986, p. 34).

Making a response

Once the listener has identified something she wishes to respond to, she must carry out the response. A non-response becomes extremely ambiguous; it is difficult to distinguish lack of attention from confirmation that all is well from a deliberate

snub. If she chooses to make a positive response, she must choose her media. Here she faces the same problems coordinating speech and text that the speaker faced in making the contribution. However, the problems play out a bit differently.

If she responds verbally to a verbal presentation, all is well. If she responds verbally to a textual item, not using the tool at all, she may have to do more work, because she does not know that other people have received the item and consider it an active element in the conversation. She may have to establish what she is talking about, for example, by reading the item aloud.

If a listener responds in text to either a verbal or a textual contribution, then there is a considerable chance for her act or the meaning of her act to be lost. For example, a person could acknowledge agreement by building upon one idea with a related idea or by writing down the idea just mentioned. Alternatively, the same act of beginning to write could signify an attempt to propose a new topic. The lack of coordination between the contribution and the response means neither the fact that one is making a response nor the particular meaning of the response will be fully available to others. In the absence of speech, even if the others notice that the responder has begun to type, there will be nothing to mark her action as a response to a particular contribution. Furthermore, if the recipients are not monitoring exactly the right space on their screen (lack of sequentiality) some unpredictable amount of time after she finishes (unpredictable delay), they may well fail to register that any change was made at all. If the change is made in an annotation, then not only must they each wait for it but they must perform an extra action to find it in the text of the item (short labels). It is quite possible for someone to re-open something that they typed originally and find that the annotations are quite different and no longer make sense to them.[†]

Just like the speaker, the recipient can also hope to improve on this difficult situation by including both written and verbal components in her response. However, she cannot just say “yes” and begin to write, but must say “yes, I’ll write that down” or “yes, even better” to let the other’s know the meaning of her beginning to type at that instance. Even if she does this extra work, the others will not necessarily know that what she has written down is consistent with her “yes” until they see it. (She could also tell them what she is writing, but since she is typing, she is in no position to pursue subsequent discussion. This behavior would represent yet more work that had to be done just to manage the logistics of the technology.) By the time she has finished typing, they may well have moved on to other things and the value of the response built on what was said, making a bid for the floor, or in any way guiding the discussion, is lost.

While the burden on the recipient is increased, her ability to respond effectively is curtailed. One way she could handle this increased burden is by inserting some sort of meta-comment, such as a question about why the item looks the way it does, into the text of the item. However, this behavior has strong consequences. For the query to carry enough information for whoever in turn receives it to know how to answer in the absence of sequential order, whatever was typed has to obey the principle of distant responsibility. That means that the responder has to take the time to anticipate what others in the future will need to see to be reminded of the issue. To

[†] Someone may well open an item just to see the annotations and not as a conversational act at all. The mere fact that there is activity does not signify change.

make such a response, the responder might well have to drop out of the conversation. If people drop out of the conversation too much, it is no longer a conversation, and they are no longer working collaboratively. Indeed, we may speculate that something like this is what happened in our first group in which each person worked separately until they turned to pencil and paper for the work they really needed to do together.

Thus, there are difficulties in both making contributions and in making responses. The contribution phase could probably be handled. However, the burden on the listener seems to be too great. Our users complained that they couldn't see enough. Our interpretation of those complaints is that users probably could have waited for the representational content of the items. They did this in the one case that seemed unproblematic for the group, when the text was reintroduced by being read aloud. What they needed to "see" immediately was the written component coordinating the interaction. The absence of coordination between speech and writing led to a highly frustrating experience.

CO-REFERENCE

Clark and Wilkes-Gibbs (1986) point out that shared reference is something that must be achieved. We had hoped that the differences between the displays at Cognoter stations would be transparent to the users or at least be accepted without too much difficulty. However, in fact people were not looking at the same surface (separate screens). This meant that they lost most of the gaze and gesture information and their base-level knowledge of what was being talked about was less than with other representational media. These losses would probably not be crippling since people manage to talk and even work (Bly, 1988) over the phone. Nonetheless, built on top of this basic situation are several discontinuities which combine to make effective co-reference difficult.

Trying to find previously entered items is potentially difficult. By design, windows may be in different places and have different shapes and sizes on different people's screens (tailorable windows). Indexical descriptions such as "It's in the upper left" do not work under these circumstances. To locate an item, you must first determine that you and your colleagues are referring to the same window, then locate the item within the window. Since windows cannot be identified by position or shape, the user must fall back on searching for their titles. Windows may be moved very quickly and invisibly. Therefore, this already time-consuming work may have to be performed repeatedly.

Additionally, trying to keep track of changes to the items presents new difficulties. First, moving or substantially changing items can be done very abruptly. When a user clicks on an item icon to move it, the item greys out on other people's screens, just as it does when an item is being edited. When the user puts the item down elsewhere, it suddenly disappears from its original location on everyone else's screen and reappears somewhere else. Even if one is looking for the flash of its reappearance, it can be hard to find on a busy screen. Furthermore, given the privacy of moving and editing, it is not very interesting to look at the screen while waiting to spot something new. Lastly, unlike with traditional media, an item is not necessarily fully displayed. Therefore, someone may change the text shown in the

icon. Subsequently, another person might want to look at the annotations attached to the item and not be able to find the item because the icon has changed.

In the case of other technologies comparable to Cognoter, we have seen that position is treated as a highly salient quality and that compact reference is desirable. In Cognoter the ability to use positional information to achieve compact reference is jeopardized in part by the new facilities we are providing and in part because we provided them in a way that undermines the work that people have to do to be able to refer to positioned items efficiently.

The combination of the eight design decisions mentioned at the beginning of this section with the conversational processes involved in using the tool, made Cognoter items more difficult both to create and to use than whiteboard objects. Our users could not "see" vital portions of the conversation and they could not make sure that each one was seeing the same objects in their representational world. Although people are in general good at compensating, there are limits to their abilities and willingness to do so. Cognoter posed too hard a communicational puzzle for our users in a conversational, time-constrained situation.

Consequences for Cognoter

An important outcome of this work was the development of a second generation tool, Cnoter. The redesign was oriented towards improving the system without giving up those features, such as the ability to move and edit items, that seem to be major benefits of computation for this kind of application. As shown in Table 1, the redesign attempted to fix four out of the eight problems we have discussed.

The major problems we tried to fix were those of private editing, private moving, unpredictable delay, and tailorable windows. We introduced shared editing and moving facilities, a significant speed-up of communication between machines, and consistent positioning of windows across machines. When someone creates a new item in Cnoter, they still open an edit window, but that window opens on everyone else's screen as well. Updates are broadcast every second. Only one person can type in a shared edit window at a time; however, control can be transferred from one person to another by clicking the mouse in the window. Thus, everyone has visual access to items while they are being created, and they can even contribute actively to that creation when appropriate. Likewise, the activity of moving icons is

TABLE 1
Summary of changes motivated by the extended interactive model

Cognoter design issue	Cnoter status
Separate screens	Unchanged
Short labels	Unchanged
Sequentiality	Unchanged
Anonymity	Unchanged
Private editing	Shared editing
Unpredictable delay	Sped up communication
Private moving	Shared moving
Tailorable windows	Windows same on all screens

broadcast several times a second, so that icons no longer disappear from one place and appear suddenly somewhere else.[†] We not only sped up communication between the machines, but added some more checking to balance out the times at which the text arrived at different machines. We eliminated the occasional very long delay for information to get from machine to machine. Thus, users should not be confused by having a particular text object and not knowing whether or not others can see it. Lastly, we made window position and shape consistent across all machines. This means that people can rely on positional information for resolving the uncertainties of reference.

We have not changed our decisions about short labels, lack of sequentiality and anonymity. With the addition of shared editing and moving, the short labels found in the idea organization windows carry much less communicative weight. They seem to be adequate for the purpose of reminding users of item contents.

The lack of sequentiality interacted poorly with attempts both to coordinate the conversation and to keep track of where items were; however, whiteboards and large sheets of paper function well without maintaining explicit evidence of sequentiality. This suggests that we should distinguish between process and product when considering our design. We had failed to give users sufficient evidence of sequentiality during the process of creating items; however, we did not need to preserve this information after the fact, when the textual contributions became simple objects in the world. The addition of shared editing and moving probably provides enough sequence information, and we do not need to change the basic whiteboard metaphor.

The issue of anonymity is similar to that of sequentiality in that one can have anonymity in the process of creation and anonymity of the eventual product. As we have pointed out, the issue of who has written text on the whiteboard is of primary importance during the process of writing, when it is a component in the conversation. There is not usually a need for a permanent label of authorship. With shared editing, anonymity no longer really exists for groups our size. If people are watching the process of creation, they will see who is doing what. This can be considered a loss, since anonymity in creating an item might permit shy or low status people to make more contributions. However, to implement process anonymity viably, we would need to move away from a basically conversational paradigm.

One of the chief benefits of the forgoing analysis is that it limits what we perceive as the sources of user difficulties. Users had problems that interacted with our decisions about sequentiality and anonymity. However, we came to believe that the designs we had in these areas were not themselves the cause of the problems. This thinking also applies to what is from some perspectives the fundamental problem with Cognoter: that we give users separate screens.

The thought that separate screens are the root of the problem is a very serious one, since they were a premise of the original system and the multiple arrangement is one of the most promising features of Colab. It is extremely flexible and presents the possibility of moving between private and public work. Furthermore, with the multiple screen/multiple keyboard setup, people can type at the same time, which means that, as with a whiteboard, the controls over whether they choose to do so are social and decided upon in response to the situation.

[†] We believe that this improvement allows perceptual rather than cognitive processing of positional changes, as described in Robertson, Card & MacKinlay (1989).

Although the analysis we have presented is by no means exhaustive, having a rich account of how the users probably came to have the difficulties they did, gives us reason to believe that the difficulties were not inherent in the multiple screen arrangement, but had limited, identifiable causes in what information was displayed on the screens. This means that, rather than giving up on the whole idea, we can try to address the problems in a more local fashion.

The new system has yet to be tested and evaluated. However, preliminary use by non-members of the Colab project suggests that users are much happier. At least one group has been able to work in the system using the computational shared workspace without the distress that we saw in our previous groups.

Other consequences

These changes fix the immediate problem. Since the Colab project's ultimate interest is in higher-level processes such as brainstorming and argumentation built on top of this fundamental conversational level, we needed to fix other difficulties as quickly as possible. Getting conversation right enables us to ask the questions about working with the system that we started with. For example, do the facilities we provide decrease "production blocking" and enable more efficient and effective idea generation and development? Do shy people contribute more or less when given a keyboard? Do people actually have a greater shared understanding of the material due to the less cryptic notes that get taken?

However, in our solutions, we tried not to experiment but simply to provide the facilities to enable successful conversation. The conversational mechanisms themselves provide another interesting source of technological experimentation and possibility. For example, although we now coordinate each participants' window size and shape, and make editor windows as well as item organization windows public, it would be interesting to see what happened if we made this user-tailorable. We might find out under what circumstances it is necessary or desirable for everyone to see the same things in the same way. We might also find out whether there are a class of situations in which it is not necessary for everyone to see the same things in the same way.

Another issue is that we have made Cognoter more conversational, rather than less. However, there is nothing about the Cognoter technology that intrinsically demands that it be used for closely coordinated work. In our studies, we set up a situation in which people would use the technology conversationally. We asked them to bring a task that they wanted or needed to work on together in the course of their own work. Presumably there are reasons why these tasks needed group solution. Our description of the technology revolved around its cooperative nature and its benefits for collaboration. Users were driven by the task, by our set-up, and by their own expectations to attempt to work by minimizing the least collaborative effort and by enjoying the benefits of mid-course correction and listener response. This must always be the case when people are actually working together. However, instead of trying to support activities already done in meetings, we could have tried to make it easier to bring activities normally done in isolation at least partially into a public forum, as when one wanders down the hall and asks a colleague about a word or a paragraph, or the outline of an algorithm, or the arguments to a function without actually involving them in the central work that is being done. The challenge here is

to provide easy flow back and forth between the public and the private at the right grain-size. This requires strategies for getting people's attention and agreement about moving from private work to collaboration. These strategies, whether provided through the computer or verbally, must be conversationally viable.

This analysis also has benefits in terms of the study of conversation and interaction. Several questions have been raised: one is how a more complete description of the relationship between talk and action in traditional representational media, might look. What are the mechanisms and resources available to people for incorporating written items into the conversation? Does some more subtle kind of try-marker occur in whiteboard use? Are there certain categories of communications that are never written on whiteboards? Why do we ordinarily not see marks like "?????" on whiteboards? Some of these questions have technical implications. For example, would it be helpful to render our notions of certainty about particular items visible in some fashion?

Conclusions

We realize that there were aspects of the original Cognoter system which did not work, and we have given some account of why. We claim that many of the serious problems in Cognoter stem from a culturally prevalent, easy-to-make assumption that communication consists of bits of verbal or textual material passed whole from one person to the next. Under this model, messages could be created, packaged, and sent by one person, unpacked, interpreted, changed, and sent on by others without regard for the exact moments of their creation and distribution. We also believe that we underestimated the problem for users of determining shared reference and therefore allowed too much activity that undermined its reliability. These underlying assumptions permitted a small but crucial number of design decisions in Cognoter that were responsible for much of the difficulty.

While it is possible that given enough time people could, as we did ourselves, learn to work with the system better than our users did, the initial difficulty was too great for a useful tool. Furthermore, whatever efficiency we gained by avoiding "production blocking" through parallel typing, this was certainly lost in the increased amount of work that had to be done to maintain communication.

We had reason to continue work on Cognoter. For one thing, we learned that certain aspects of the system were positively received. Promisingly, the group that used the video network to work around the problems discussed, was quite happy with the system. Although they used only a fraction of its functionality, some felt it could become "addictive", and commented with enthusiasm on the ease of bringing a fourth colleague up to speed on the work they had done in their Cognoter sessions.

Furthermore, we have yet to evaluate the features that we imagine as the most important contributions of the tool. By and large, the qualities we sought to promote in patterns of work are untouched by this analysis. Brainstorming, organizing and evaluating is still an interesting process to support. The ability to use the keyboard, the ability to save, printout, and recall organizations, the ability to rearrange material, and to handle large amounts of material are all important features.

However, we now see that the process we first identified at the whiteboard and later brought to the computer was built on conversational abilities. This affects some of our ideas. For example, we promoted the notion that it would be more efficient to work in parallel. That idea remains; however, it now rests on a deeper notion of what working in parallel means. It does not mean merely working at the same time as someone else, which is after all, what we do when we work alone. Instead, it means giving participants the ability to judge when it is appropriate to overlap, just as they judge the efficacy of other possible moves in a conversation. In other words, to work successfully in parallel, we must have resources for working together.

What happened with Cognoter has significance for CSCW beyond the boundaries of this particular project because Colab was a highly innovative project that encountered serious difficulty because it did not recognize that it had entered a new arena. It slipped up on implicit aspects of the system, places where the system designers didn't realize they were making choices. Although these problems were suspected in preliminary observations, two factors common to CSCW research conspired to make them difficult to pin down convincingly. One is that the problems are at such a fine level of coordination that is not possible to capture and recreate specific incidents of problems without electronic means. The second is that considerable knowledge of the system and its rationale does seem to allow people to function without overt distress, perhaps by causing them to reduce their criteria for understanding what is happening at particular moments.

Human-machine studies are done typically to examine those aspects of systems that are already deemed important for success. In a field that is as new and as complex as computer-supported cooperative work, and equally as liable to fail (Grudin, 1988), highly directed studies need to be augmented by other approaches such as undirected observation. Furthermore, system designers must draw on and reason about social science results such as the interactive model that are not necessarily predictive and which do not necessarily describe the exact situation that they are designing. This is a risky but potentially rewarding strategy.

The ability to use whatever social science insights we have to feed into careful thought about the situation we are trying to create is crucial because, as designers of novel technologies, we must judge whether the technologies we envisage are likely to work. While it is easy to judge failure by the distress expressed by users, it is much harder to judge the success and potential of a system that has encountered difficulty. Although this analysis has focused on problems and addressed potential only incidentally, it also suggests that Cognoter was close to being able to create a multi-user system that people would be able to use with the ease, range of expressive behaviors, and mastery that they employ in conversation. Furthermore, experience suggests that perhaps we can give users much more facility due to the potentially greater expressive power of the computerized representational medium.

Many thanks are due to Sara Bly, Frank Halasz, Steve Harrison, Leigh Klotz, Cathy Marshall, and John Tang for reading early drafts and giving excellent comments. Robert Anderson, Liam Bannon, Marjorie Goodwin, Jonathan Grudin, Susan Newman, Wendy Mackay, Lucy Suchman, and Mark Stefik all contributed through comments and extensive discussion. As usual, Stan Lanning was indispensable in helping us implement the software. Additional thanks to the anonymous reviewers. We were grateful for their suggestions and hope they are pleased with the results.

References

- BLY, S. (1988). A use of drawing surfaces in different collaborative settings. In *Proceedings of the Conference on Computer-Supported Cooperative Work*, pp. 250–256, Portland, Oregon, September 26–28.
- CLARK, H. H. & SCHAEFER, E. F. (1987). Collaborating on contributions to conversations. *Language and Cognitive Processes*, **2**, 19–41.
- 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.
- DIEHL, M. & STROEBE, W. (1987). Productivity loss in brainstorming groups: toward the solution of a riddle. *Journal of Personality and Social Psychology*, **53**, 497–509.
- DUNCAN, S. D. (1973). Toward a grammar for dyadic conversation. *Semiotica*, **9**, 29–47.
- FOSTER, G. (1986). Collaborative systems and multiuser interfaces. PhD. thesis, University of California, Berkeley.
- FOSTER, G. & STEFIK, M. (1986). Cognoter, theory and practice of a Colab-orative tool. In *Proceedings of the Conference on Computer-Supported Cooperative Work*, pp. 7–12. Austin, Texas, December 3–5.
- GOODWIN, C. (1981). *Conversational Organization: Interaction Between Speakers and Hearers*. New York: Academic Press.
- GOODWIN, C. & HERITAGE, J. (1990). Conversation analysis. *Annual Review of Anthropology*, **19**, 283–307.
- GOODWIN, C. & GOODWIN, M. H. (1990). Context, activity, and participation. In P. AUER & A. DI LUZO, Eds. *The Contextualization of Language*. Amsterdam: Benjamins.
- GOODWIN, M. H. (1980). Processes of mutual monitoring implicated in the production of description sequences. *Sociological Inquiry*, **50**, 303–317.
- GRUDIN, J. (1988). Why CSCW system fail. In *Proceedings of the Conference on Computer-Supported Cooperative Work*, pp. 85–93, Portland, Oregon.
- POMERANTZ, A. (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. New York: Academic Press.
- POMERANTZ, A. (1984). Pursuing a response. In J. M. ATKINSON & J. HERITAGE, Eds. *Structures of Social Action*, pp. 152–164. Cambridge, UK: Cambridge University Press.
- ROBERTSON, G. G., CARD, S. K. & MACKINLAY, J. (1989). *The Cognitive Co-processor Architecture for the Interactive User Interfaces*, Xerox PARC Technical Report #SSL-89-28.
- SACKS, H. & SCHEGLOFF, E. A. (1979). Two preferences in the organization of reference to persons in conversation and their interaction. In G. PSATHAS, Ed. *Everyday Language: Studies in Ethnomethodology*. New York: Irvington.
- SCHEGLOFF, E. A. (1981). Discourse as an interactive achievement: some uses of uh-huh and other things that come between sentences. In D. TANNEN, Ed. *Analyzing Discourse: Text and Talk*. Georgetown University Roundtable on Languages and Linguistics 1981, pp. 71–93. Washington, DC: Georgetown University Press.
- SCHEGLOFF, E. A. & SACKS, H. (1973). Opening up closings. *Semiotica*, **8**, 289–327.
- SCHÖBER, M. F. & CLARK, H. H. (1989). Understanding by addressees and overhearers. *Cognitive Psychology*, **21**, 211–232.
- STEFIK, M. J., FOSTER, G., BOBROW, D. G., KAHN, K., LANNING, S. & SUCHMAN, L. (1987a). Beyond the Chalkboard. *Communications of the ACM*, **30**, 32–47.
- STEFIK, M. J., BOBROW, D. G., FOSTER, G., LANNING, S. M. & TATAR, D. G. (1987b). WYSIWIS revised: early experiences with multiuser interfaces. *ACM Transactions on Office Information Systems*, **5**, 147–167.

- SUCHMAN, L. (1988). Representing practice in cognitive science. *Human Studies*, **11**, 305–325.
- TANG, J. C. (1989). *Listing, Drawing and Gesturing in Design: A Study of the Use of Shared Workspaces by Design Teams*, Xerox PARC Technical Report #SSL-89-3 (Ph.d. thesis Stanford University).
- TANG, J. C. (1990). *Observations of the Use of Shared Drawing Spaces*. Videotape, Xerox PARC Technical Memo.