

CO-ORDINATING ACTIVITY: AN ANALYSIS OF INTERACTION IN COMPUTER-SUPPORTED CO-OPERATIVE WORK

Steve Whittaker

Susan E. Brennan

Herbert H. Clark

Hewlett Packard Laboratories, Bristol
& HP Stanford Science Centre
& Psychology Department
Stanford University
sjw@hplb.hpl.hp.com

Psychology Department
SUNY at Stony Brook
NY 11794
brennan@psych.stanford.edu

Psychology Department
Stanford University
herb@psych.stanford.edu

Abstract

We examined mediated communication using a shared electronic Whiteboard *with* and *without* the addition of a speech channel. The 3 users were not co-present. There were two major findings: (a) permanent media such as the Whiteboard enable users to construct shared data structures around which to organise their activity, and (b) this permanence allows users to abandon some (but not all) of the turn-taking commonly used in spoken conversation and to organise their activities in a highly parallel manner. With the addition of a speech channel, people still used the Whiteboard to construct shared data structures that make up the **CONTENT** of these communications, while speech was used for coordinating the **PROCESS** of communication.

KEYWORDS

Mediated communication, group work, media, shared workspaces, activity co-ordination

INTRODUCTION

There has been much recent interest in systems that support real-time group activity for both office tasks and multi-participant design [Bly88, GS88, SFB⁺88, TL88]. These systems attempt to support communication between two or more people who are usually not physically co-present, by providing an array of different communication media. Media can include a shared writing or sketching space, audio, video, or combinations of these. The promise of these systems is to enable collaborations to occur between distributed workgroups, or support group communication by supplying tools specifically designed to facilitate the different phases of meetings [Bly88, SFB⁺88]. They have not been greatly successful however [Gru88, TFB91]. Why should this be, when the technology of text, audio and video are now readily available to support and augment communication? Our claim here is that CSCW systems have been designed in the absence of a theory of how people communicate, and that such a theory is needed to produce appropriate combinations of electronic media to support communication.

In what ways is mediated communication likely to differ from face-to-face interaction? Previous work has compared the effects of using different media on communication. Although it documents the efficiency of speech[COPW72, Cha75, SWC76] it does not offer explanations for its utility, nor does it explain why adding other media to speech fails to improve communication. Elsewhere we have identified a set of properties of different media and argued how they influence communication[CB91]. We looked at one media property here, that of permanence, to investigate how it affects communication. In most mediated interactions, there are *permanent* media such as video, typed text and drawings, as well as *ephemeral* media such as speech and gesture. Previous work has begun to address this difference between permanent media and ephemeral conversation [Bly88, TL88].

We set out to examine the effect of permanence. In our first study, we looked at interaction when the sole means of communication was by typing, writing or drawing on a shared electronic Whiteboard. Participants were at different physical locations, so they could not communicate using speech or gesture. We predicted that communication using this exclusively permanent medium would differ from face-to-face communication in several ways. First, permanent media may not require the serial unfolding of topics that characterises speech[Lev83]: Contributions persist, so a participant need not reply to another person's input immediately, because they know that the input will not disappear. This should lead to more parallel activity with permanent media. Second, in speech, relations between utterances are signalled by temporal contiguity [Gri75, SSJ74], whereas in permanent media, other devices such as spatial placement are available for signalling these relations. Temporal contiguity may only be such a strong cue in speech because of the need to reply to an utterance immediately, before its content is forgotten. Thirdly, permanent electronic media give rise to a record of the interaction which can serve as focus and reference point for coordinating the group's activities: this record can be invoked for reference and also manipulated and incorporated with external materials. A final difference between the media types is that permanence raises problems that do not occur with ephemeral media, namely that participants have to manage their use of space in the medium because space is a limited resource.

We therefore set out to test these predictions. The purpose of the first study is to characterise communication using a permanent medium, and to explain the differences between this and ephemeral conversation. In a second study we examine the effects of adding an ephemeral medium to the Whiteboard in the form of a speech channel. The question we address is which communicative functions are achieved in the ephemeral and which in the permanent medium when people have a choice of media types? The object is to use these observations to construct a theory of mediated communication.

METHOD: THE SYSTEM AND THE TASKS

We looked at 6 groups of 3 people who collaborated using the "XScrawl" Shared Whiteboard running on workstations. (See Figure 1). Participants could type or draw inputs wherever they placed their cursors¹. Cursor placement was done with a mouse and transmission of information was instantaneous. Drawing was done with the mouse, and users chose to type or draw by selecting the appropriate menu option. To identify who was presenting what input to the Whiteboard, each user's inputs appeared in a different colour². There were no constraints on simultaneous input, so all three participants could produce input at once. Participants' cursors were not visible to one another: the system had a menu option to allow publicly visible pointing. The size of the window was 55 by 28 typed characters: because of the permanence of this medium and the limited screen size, material had to be deleted often. It could be deleted either using a selectable eraser or using the "clear screen" menu option. There were no restrictions on what could be deleted, so people could delete material created by someone else. The 3 users all worked at different physical locations in Hewlett Packard Labs, and the Whiteboard was their only means of communication.

After a brief tutorial on the system, users had several minutes to familiarise themselves with it, before working together on 2 tasks³. The first was a brainstorming task in which the objective was to produce a prioritised list of 7-10 items to complete the following sentence: "When buying a house the critical features that the house must possess are ...". The second involved calendar coordination: each participant had some but not all of the information necessary to complete the task: to arrange two two-hour meetings that all three could attend. We videotaped both tasks and conducted the analysis on the videotapes.

THE DATA AND CODING SCHEME

Participants did not proceed by typing textual messages to one another in strict sequence. First, not all the inputs were textual. People exploited the fact that they could place inputs anywhere on the screen to generate lists of items, tables, calendars and matrices for voting and rating. They used drawing and pointing

¹This is different from many of the studies of media and communication which have employed a console input[COPW72, Cha75, OC89b, OC89a]. Console input both imposes a specific sequential spatial organisation on the dialogues, and it also means that material is ephemeral as it can scroll off the top of the screen.

²The colour cue is of course absent from the Figure; we have reconstructed the inputs in 3 different fonts.

³These were similar to those used by Gale [Gal89].

to indicate relations between inputs, to select, to mark, and to draw attention to particular objects. Our hypothesis was that LOCATION in permanent media would serve some of the same functions that temporal contiguity does in face-to-face interaction. Therefore both LOCATION and TEMPORAL properties of inputs were coded.

An input was defined as any sequence of typed characters or drawn marks generated by one person. The sequence could not contain pauses of more than one second. Inputs ranged from a single drawn mark or character, to several sentences of typed text⁴. Inputs fell into three distinct classes, namely (a) ARTIFACTS(30% of inputs): Lists, tables, calendars, matrices, including the drawing of lines and tables and the material that was inserted into these. A complex ARTIFACT could be constructed in multiple inputs by several people. (b) PROSE(56% of inputs): Textual material that was not part of an ARTIFACT. (c) DEIXIS(14% of inputs): Drawing, pointing, selecting and voting when this involved reference to previously existing material. Voting was defined as placing a numeric rating or score next to a previous item or phrase⁵.

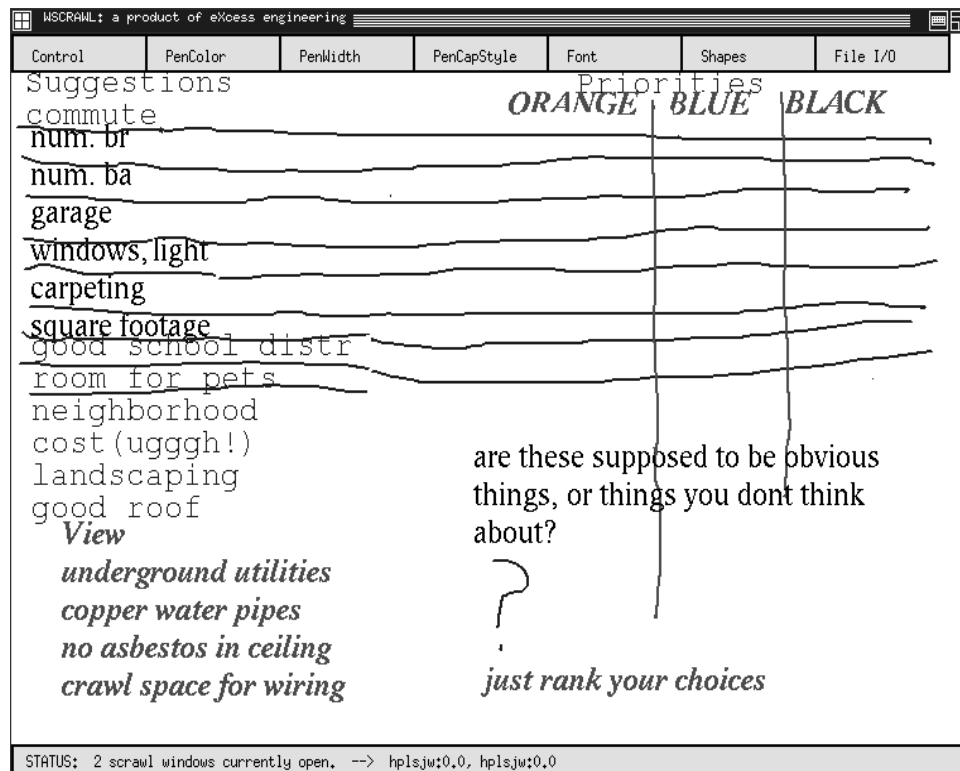


Figure 1: Screen Dump for House Selection Task

Figure 1 occurred halfway through the brainstorming task. Inputs by the different users appear in different fonts. On the left side is a vertical list of suggested items constructed by all users, and on the upper right is a table with columns for each person for rating the priorities of the items(ARTIFACT). The lower right contains a textual question concerning the status of items on the list (PROSE), a query mark (DEIXIS), and a textual answer to the question (PROSE).

People used LOCATION as one means to indicate relations between inputs. Proximity judgments were made from the physical centre of each input by a coder who was blind to the hypotheses of the experiment. The aim was to determine which of the previous inputs the current input was nearest to. We identified three different placement strategies based on proximity and timing information: (a) PREVIOUS: The current input was placed nearest the previous input that had been produced regardless of who produced it. This occurred when all the participants worked in the same screen area with each person placing their current input adjacent to the previous input. (b) SELF: The current input was placed nearest to another input

⁴The groups successfully completed each task in times ranging from 5.5 to 31.6 minutes with a mean of 16.6 minutes and 135 inputs. There were 1619 inputs in total.

⁵To check reliability, 327 inputs (20%) were independently coded by the first author and a coder who was blind to the experimental hypotheses. Reliability was 0.95.

by the same person. For example, users defined specific regions for the exclusive use of each person and typed their inputs into these segregated areas. (c) ELSEWHERE: All other placements excluding the above. For example, when users placed inputs in a neutral area when parts of the screen had been personalised. In Figure 1, users have defined personal areas at the top, centre and bottom of the left side(SELF). The question, the mark and the answer on the lower right are organised according to the PREVIOUS placement strategy.

For the purely TEMPORAL properties, the onset time, the completion time and the DURATION of an input on the screen were coded. DURATION was defined as the total time that an input remained on the screen after it was completed. We also noted the OVERLAP with the production of other inputs, i.e. whether an input was produced at the same time as another input was being produced. Finally, the coder noted whether people consulted others before deleting a particular input, CONSENSUAL DELETE.

RESULTS

As predicted we found that using a wholly permanent medium produced large amounts of parallel activity with 41% of inputs being produced at the same time as some other input. This contrasts sharply with spoken conversation, where overlapping is typically 5% or less [Lev83]. This parallelism meant that participants could no longer rely on temporal contiguity as the main indicator of relations between inputs. This weakening of the power of the temporal cue was shown by the fact that even when participants were engaged in sequential non-overlapping activity, they often supplemented the timing cue by placing their current input **next** to the previous input. We have already noted the use of LOCATION to indicate other types of relations between inputs: some groups personalised specific areas of the Whiteboard for the exclusive use of each user. Personalisation was often combined with a group area in which all users placed their inputs.

While these overall results support our initial predictions concerning interaction in a permanent medium, what was more striking was that the different input types, ARTIFACTS, PROSE and DEIXIS were radically different as far as timing, location and deletion regimes were concerned. These differences seemed to relate to the fact that the input types were directed towards different communicative functions.

We first compared PROSE and ARTIFACTS and found that ARTIFACTS lasted FIVE times as long as PROSE ($t_{(11)} = 25.34$)⁶. and also that people were more likely to seek permission to delete an ARTIFACT than PROSE ($t_{(11)} = 3.94$)

How should we interpret these differences? It seems that ARTIFACTS such as lists, tables, calendars and matrices are being used to carry the CONTENT of communication. They are used as shared datastructures which are referred to, pointed at, marked, and rated in the course of constructing a solution. In contrast, PROSE is being used for PROCESSES such as discussion, clarification, comment and consensus about that CONTENT. For example a person might use PROSE to clarify an item in a list. By this analysis, we should expect ARTIFACTS to last for much of the task. PROSE, however, can be deleted once its purpose has been achieved and its preservation is no longer critical to the task. By the same argument we should expect people to be much more circumspect about the deletion of ARTIFACTS than PROSE because of their information bearing function and the fact that they are joint products. They should therefore be more likely to consult the other members of their group before deleting an ARTIFACT than a piece of PROSE (See Figure 2).

		Prose	Artifact	Deixis
Temporal	Duration (secs)	66.7	322.2	158.8
	% Overlap	38	51	29
Spatial	Previous	27	10	4
	Self	44	66	36
	Elsewhere	30	24	60
Deletion	% Consensual Delete	46	72	53

Figure 2: Comparison of input types

We also found that PROSE and ARTIFACTS showed different spatial placement and sequential co-ordination. A critical element of PROCESSES like clarification and negotiation is that they are dependent on a high degree of temporal co-ordination between participants. Thus if one participant wishes to achieve consensus

⁶Unless otherwise stated, all differences that are described are significant at the 0.05 level for the appropriate statistical test.

to erase an area it is crucial that s/he obtain agreement from other participants before proceeding. The same is true if one person requests information from another; they sometimes cannot proceed until it is supplied. Attention is vital: participants must not “overlook” or “miss” attempts to engage in PROCESSES, particularly if another person requires a response before they can continue. We might therefore expect that when people are engaged in PROCESSES that they are more likely to all focus their activities in the same area of the screen. In contrast, with ARTIFACTS there is not as much need for close sequential or spatial co-ordination; when people are supplying CONTENT in an ARTIFACT their actions can be much more parallel because those actions are not critically dependent on each other. Thus if all 3 participants are completing lists of their free times, they can all focus on different parts of the calendar at the same time, because there is no direct dependency between their acts.

Consistent with this we found that ARTIFACTS showed many more overlaps than PROSE and they were also more likely to be constructed in personal areas, because their construction does not require close sequential co-ordination (See Figure 2). The opposite is true for PROSE: participants were more likely to all focus on the same screen area because it is imperative that their activities be closely coupled and that messages are not overlooked. (PROSE more likely than ARTIFACTS to be next to previous input, $t_{(11)} = 4.81$, ARTIFACTS more likely than PROSE to be in personal locations, $t_{(11)} = 3.28$). Problems of co-ordination arose when people attempted to engage in PROCESSES, while either using different areas of the screen or unpredictable spatial placements: We observed several examples of people temporarily abandoning their own private area and retyping an input in someone else’s area, when their first input had received no response.

A major advantage of using an ARTIFACT is that it allows rapid parallel activity for the construction of shared datastructures. We should therefore expect groups who made more extensive use of ARTIFACTS to be more efficient in completing their tasks and indeed we found that groups who produced proportionally more ARTIFACTS were faster to complete the tasks overall (Pearson $r_{(11)} = 0.55$, $p = 0.07$). This result is striking given that there is an overhead associated with the construction of ARTIFACTS: people have to construct a matrix or table before they can fill it in.

What about the remaining actions such as pointing, drawing and voting? These are used to draw attention to, or mark existing material. It seemed that DEIXIS occurred once people had a shared datastructure such as an ARTIFACT to serve as the focus for activity. Deixis was more likely to refer to ARTIFACTS than PROSE ($t_{(11)} = 2.45$). As far as spatial placement is concerned, we should expect DEIXIS to often be located “Elsewhere” because drawing or voting can be applied to any previous material, not just one’s own or recently constructed material (See Figure 2). We made no predictions about the temporal behaviour of DEIXIS. On some occasions, DEIXIS was handled in a sequential manner, e.g. one person might give the answer to a question by pointing to an item. On others however it was highly parallel as when people searched through 3 lists and placed a rating of 1-10 by the items of their choice.

ADDING A SPEECH CHANNEL

Other research indicates that speech is the most effective single communication channel and that adding other media does not significantly improve communication for information retrieval tasks [Cha75, OC89b]. Our findings show, however, that ARTIFACTS serve a critical function as a shared datastructure, so we should still expect groups to construct ARTIFACTS even when a speech channel is available. We expected this even though ARTIFACTS are more time-consuming to produce than speech because they have to be typed or drawn.

We added a speech channel to the shared Whiteboard, in the form of a three-way telephone conference call, with telephone headsets for hands-free operation. We had a second set of six groups of three subjects do the the same two tasks. On 10 of the 12 occasions those groups still produced ARTIFACTS, indicating the utility of such structures. Indeed we found that the number of inputs that were classified as ARTIFACTS were comparable across the two experiments (486 in the first study and 556 in the second). We found strong independent evidence for our analysis of PROSE as PROCESS. Given that PROCESS both has to be tightly co-ordinated and does not need to persist once its immediate purpose has been achieved, we should expect almost no typing of PROSE because all PROCESS should take place in the speech channel. This is what we found: Only one group produced more than 3 PROSE inputs and altogether PROSE accounted for only 1% of inputs. Surprisingly, however, the amount of DEIXIS also decreased with only 30 instances being observed compared with 227 in the first study. This was because people did not vote on the Whiteboard by

placing ratings by the items of their choice as they had in the first study. Instead they verbally negotiated the relative importance of various items. It seems that the voting by DEIXIS in the first experiment was a strategy that subjects developed to overcome the inherent unsuitability of the Whiteboard for activities like negotiation. A final difference in the second study was that we observed activities like doodling. These may only occur with speech because people can still contribute verbally to the interaction while doodling on the Whiteboard. In the first study, doodling prevented other forms of interaction. In addition doodles may have been interpreted as serious contributions.

CONCLUSIONS

What are the implications of this study for the design of CSCW systems? First, given the effectiveness of ARTIFACTS it would seem that these could usefully be incorporated into CSCW systems as permanent tools. Thus a system could have calendars, matrices, and lists as well as other application-specific ARTIFACTS which could be invoked if the task demanded it. Consensus would be necessary for erasing ARTIFACTS to prevent deletion of shared datastructures. In fact organisations already make use of ARTIFACTS such as spreadsheets as a focus for co-ordinating financial activity such as planning or targetting. Systems based around these ARTIFACTS have been built for single users interacting with a system [SWH⁺90, NM90].

The second implication concerns the organisation of initiative and the function of different media in this type of system. Designers have made decisions about whether or not to include a turn-taking regime in their system. What the current study shows is that the usefulness of turn-taking depends on the specific activity the users are currently engaged in. ARTIFACT construction was a highly parallel activity in contrast to PROSE which was generally more serial. The optimal system would therefore enable users to switch between regimes depending on their current activity. One method for supporting this would be to match the properties of the media to the local demands of the communication [Wal89, WW89], so that speech could be used for clarification, negotiation, and question answering, whereas typing or drawing could be used for the representation and exchange of complex data. When multiple media are not available, however, it may be necessary to allow different turn-taking regimes for different activities within a medium. Finally there are both advantages and disadvantages for different media. While finding speech highly efficient in most of the second study, our users commented that for certain types of meeting, a permanent record of PROCESS would be useful. Elsewhere we have outlined a theoretically motivated analysis of the costs and benefits of different media in communication [CB91].

We have also noted that communication using permanent media differs from face-to-face conversation and we were able to predict these differences. Permanence affords a different style of parallel interaction supported by ARTIFACTS which does not obey the rules of turn-taking: The parallel construction of these complex shared datastructures is made possible because the permanent medium frees participants from the incremental discussion of data which would be necessary in verbal communication. This function of permanent media has been overlooked in previous work which has argued for the inherent superiority of the speech channel [Cha75, OC89b]. Our finding in the second study, that participants still chose to produce typed or drawn ARTIFACTS, even when they had the speech channel, shows the utility of channels other than speech. It may be that the real benefits of ARTIFACTS only begin to emerge when there is highly complex information to be shared. Shared datastructures also serve as the focus for DEIXIS. Once they have ARTIFACTS, participants can identify, mark, manipulate and isolate key elements of a complex structure using simple gestures rather than complex locutions. We observed this with the voting strategy in the first study. This supports the findings of Bly [Bly88] that users spend over 60% of the time referring to previous inputs in face-to-face design activity. We have also demonstrated the utility of such datastructures for DEIXIS in the context of communication between single users and computers [WW89, WS89].

While ARTIFACTS were highly efficient for certain parts of the Task, it was clear that there were parts of the communication for which they were ineffective because those aspects are fundamentally serial. In the second study, the PROCESSES were all in the speech channel and we observed almost no typing of PROSE. For the Whiteboard alone, these PROCESSES, such as clarification and consensus, that require close sequential co-ordination, were most effectively handled when people abandoned the strategy of typing in parallel in different areas of the screen and employed a regime of turn-taking in one restricted region. Effective systems for collaborative communication will have to support these different activities by matching the properties of the media to the demands of the communication.

ACKNOWLEDGEMENTS

Thanks to Lyn Walker, Phil Stenton, David Frohlich, Bonnie Nardi, Robin Jeffries and Jim Miller for feedback on early drafts, to Steve Gale for showing us his data from a similar study, and to Beth Bryson for transcription and coding.

References

- [Bly88] Sara S. Bly. A use of drawing surfaces in different collaborative settings. In *Proceedings of the Conference on CSCW*, pages 250–256, 1988.
- [CB91] Herbert H. Clark and Susan E. Brennan. Grounding in communication. In L. B. Resnick, J. Levine, and S. D. Teasley, editors, *Perspectives on socially shared cognition*. Washington: APA Press, 1991.
- [Cha75] A. Chapanis. Interactive human communication. *Scientific American*, 232:34–42, 1975.
- [COPW72] A. Chapanis, R.B. Ochsman, R.N. Parrish, and G.D. Weeks. Studies in interactive communication: I. the effects of four communication modes on the behavior of teams during cooperative problem-solving. *Human Factors*, 14:487–509, 1972.
- [Gal89] Stephen Gale. The vision project. Technical Report HPL-BRC-TR-89, Hewlett-Packard Laboratories, Bristol, U.K., 1989.
- [Gri75] H. P. Grice. *Logic and Conversation*. Dickenson Publishing Co., 1975.
- [Gru88] Jonathan Grudin. Why csw applications fail. In *Proceedings of the Conference on CSCW*, pages 85–93, 1988.
- [GS88] Irene Greif and Sunil Sarin. Data sharing in group work. In Irene Greif, editor, *Computer-Supported Co-operative Work*. Morgan Kaufmann, San Mateo, Ca., 1988.
- [Lev83] Stephen C. Levinson. *Pragmatics*. Cambridge University Press, 1983.
- [NM90] Bonnie Nardi and Jim Miller. The spreadsheet interface. In *Proceedings of interact*, Cambridge, England, 1990.
- [OC89a] Sharon L. Oviatt and Philip R. Cohen. The contributing influence of speech and interaction on human discourse patterns. In J.W. Sullivan and S.W. Tyler, editors, *Architectures for Intelligent Interfaces: Elements and Prototypes*. Addison Wesley, Menlo Park, Ca., 1989.
- [OC89b] Sharon L. Oviatt and Philip R. Cohen. The effects of interaction on spoken discourse. In *Proc. 27th Annual Meeting of the Association of Computational Linguistics*, pages 126–134, 1989.
- [SFB+88] Mark Stefik, Gregg Foster, Daniel Bobrow, Kenneth Kahn, Stan Lanning, and Lucy Suchman. Beyond the chalkboard: Computer support for collaboration and problem solving in meetings. In Irene Greif, editor, *Computer-Supported Co-operative Work*. Morgan Kaufmann, San Mateo, Ca., 1988.
- [SSJ74] Harvey Sacks, Emmanuel Schegloff, and Gail Jefferson. A simplest systematics for the organization of turn-taking in conversation. *Language*, 50:pp. 325–345, 1974.
- [SWC76] J. Short, E. Williams, and B. Christie. *The Social Psychology of Telecommunications*. Wiley, London, 1976.
- [SWH+90] Phil Stenton, Steve Whittaker, Keith Harrison, Nick Haddock, and Andy Nelson. The gap experiment. Technical Report HPL-BRC-TR-90-157, Hewlett-Packard Laboratories, Bristol, U.K., 1990.
- [TFB91] Deborah Tatar, Gregg Foster, and Daniel Bobrow. Design for conversation: Lessons from cog-noter. *International Journal of Man-Machine Studies*, 34:185–211, 1991.

- [TL88] John Tang and Larry Leifer. A framework for understanding the workspace activity of design teams. In *Proceedings of the Conference on CSCW*, pages 244–249, 1988.
- [Wal89] Marilyn A. Walker. Natural language in a desk-top environment. In *Proceedings of HCI89, 3rd International Conference on Human-Computer Interaction, Boston, Mass*, pages 502–509, 1989. Also a Hewlett Packard Laboratories, Bristol, Technical Report.
- [WS89] Steve Whittaker and Phil Stenton. User studies and the design of natural language systems. In *Proc. 4th Conference of the European Chapter of the ACL, Association of Computational Linguistics*, pages 116–123, 1989.
- [WW89] Marilyn Walker and Steve Whittaker. When natural language is better than menus: A field study. Technical Report HPL-BRC-TR-89-020, Hewlett Packard Laboratories, Bristol, England, 1989.