# Theories and Methods in Mediated Communication

Steve Whittaker

AT&T Labs-Research

180 Park Avenue

Florham Park

New Jersey, 07932, USA

Phone: (973) 360 8339

FAX: (973) 360 8809

Email: stevew@research.att.com

.

.

.

## 1. Theories and Methods

Technology mediated communication is a fact of life. The human communication apparatus is constrained in several ways. There are limits to the distance at which speech is audible, and visible behaviours such as gesture, gaze or facial expressions are perceivable. Furthermore, these natural communication behaviours are ephemeral and do not persist over time. Given these limitations, we must rely on some form of mediation, if we are to communicate at distance and across time. People have therefore invented technologies that attempt to circumvent these limits to allow remote synchronous communication (e.g. telephone, videoconference) or asynchronous communication (e.g. letters, telegraph, email, fax, voicemail).

Understanding the principles that govern mediated communication has important practical implications. Because of the pervasiveness of mediated communication, we need to determine whether, how and why, it differs from face to face communication. If there are differences between mediated and face to face communication, we need to provide guidelines to inform decisions about the circumstances in which it is appropriate to use mediated communication, and the effects of using it in those situations. Systematic insights into mediated communication should also enable us to improve the designs of current and future technologies supporting mediated communication. An understanding of mediated communication should also inform more general theorizing about the psychology of communication. Most current communication theories regard face to face communication as an integrated set of speech, gaze, and gestural behaviours. As we shall see, studies of mediated communication allow us to isolate the individual contribution of different nonverbal behaviours, such as gaze and gesture to communication. They also help to clarify the overall influence of visual information in communication, and the effects of communication interactive processes such as feedback on communication.

The structure of the review is as follows. The main part of the review is organized around 3 main theoretical approaches to mediated communication. We first talk about the general characteristics of mediated communication theories, and methods in this area. We then characterize each theory in detail and evaluate the evidence for it. We conclude with a discussion of outstanding practical and theoretical issues.

.
.
.

Theories of mediated communication

The current section outlines: (a) mediated communication theories; (b) the technologies used to support communication in the various situations; (c) the methods used to study mediated communication. There are many different theories of mediated communication. Rather than describing each in detail, we restrict ourselves here to an exposition of the general form that these theories take, and the general set of claims that they make. The fundamental goal of mediated communication theories has been to explain the relationship between the affordances (Gibson, 1979; Norman, 1988) of different mediated technologies and the communication that results from using those technologies. Most theories are comparative, addressing how and why mediated communication is different from face to face communication. The general form of these theories is to: (a) characterize how technologies differ in terms of their communication affordances; (b) describe how the affordances of a given technology differ from those of face to face communication; and (c) explain how these different affordances produce differences between mediated and face to face communication in process, content or outcome of communication. For example, social disinhibition theory (Kiesler, Siegel & McGuire, 1984; Sproull & Kiesler, 1986) argues that email has different affordances from face to face communication. Email is asynchronous, so it cannot provide immediate interactive feedback to speakers about whether their communication has been understand or accepted. Email also does not provide non-verbal information such as head nods, gestures, facial expressions and postural information that have been argued to be important to the transmission of social and affective information (Argyle, 1990; Short, Williams & Christie, 1976). According to the theory, the absence of feedback, and the lack of social and emotional information has two effects, it: (a) inhibits social processes such as consensus formation, and (b) leads to stronger emotional expression in email.

Table 1 shows the affordances of different mediated communication technologies as identified by these theories. One major technology affordance is the different *modes* that technologies support. The main difference here is between linguistic and visual modes. For example, email, the phone and voicemail transmit purely linguistic information, whereas videoconferencing is a combination of a linguistic speech channel and a visual channel showing images of participants. Another example of a technology that

.
.
.

combines linguistic and visual information is Shared Workspaces, which combine a synchronous speech channel with shared visual material such as a documents or graphics (Tang, 1991, McCarthy, Miles & Monk 1991, McCarthy, Miles, Monk, Harrison, Dix & Wright, 1993, Stefik, Foster, Bobrow, Kahn, Lanning, S., & Suchman, 1997, Tatar, Foster & Bobrow, 1991, Whittaker, Brennan & Clark, 1991, Whittaker, Geelhoed & Robinson, 1993). Note that there are no instances of pervasive technologies that are purely visual: we discuss reasons for this in section 2. The second major affordance is *interactivity*. For example, the phone, instant messaging and chat are interactive because they support synchronous bidirectional communication, allowing participants to provide immediate feedback to speakers. In contrast, email and voicemail are non-interactive because they do not allow such concurrent feedback. Another example of a non-interactive technology is videomail, which allows people to leave messages that include both speech and pictures (Hopper, 1994). Non-interactive technologies are necessarily permanent, because message recipients are not present when the message is generated. In contrast, interactive ones can be ephemeral (e.g., the phone) or permanent (e.g., instant messaging and chat). While some theories (e.g. Clark & Brennan, 1991) propose more complex sets of affordances, modality and interactivity are sufficient to explain most of the research we report here. The conclusions section discusses the limitation of the current affordance taxonomy and describes how it might be enriched. Finally, in this review, we exclude technologies such as radio, television, newspapers and film; these technologies only allow one-way broadcast and our concern here is with two-way interaction.

| AFFORDANCE | | INTERACTIVITY | |
|---|---|---|---|
| | | INTERACTIVE | NON-INTERACTIVE |
| MODE | LINGUISTIC | Phone, Audioconference, Chat Instant Messaging | Email, answerphone, voicemail, FAX, letter, Usenet. |

.
.
.

| | | | |
|---|---|---|---|
| | LINGUISTIC AND VISUAL | Videoconference, VideoPhone, Shared Workspace | VideoMail |

Table 1 - Technologies and their affordances.

Having characterized technologies and their affordances, the theories explain how those different affordances affect communication behaviours. By determining how different affordances affect communication behaviours, mediated communication theories predict the effects of these behaviours on critical communicative phenomena involving communication content, process and outcome. The relationship between affordances, behaviours and their communicative effects are derived from studies of the role that different communication behaviours play in face to face communication. By identifying how visual behaviours and interactivity affect these behaviours, the theories make predictions about how communication processes, content and outcome will be altered in technologies that do not support visual modes or interactivity. For example, face to face research has shown how behaviours such as head nods, gaze and gesture mediate turn-taking (Argyle & Cook, 1976; Argyle, Lalljee & Cook, 1968; Argyle, Lefebvre & Cook, 1974; Beattie, 1978; 1981; Duncan, 1972; Kendon, 1967). However each of these behaviours is visible and depends on the presence of visual information. We might therefore expect that technologies that do not support visual information would show impaired turn-taking (Jackson, Anderson, McEwan & Mullin, 2000; O'Conaill, Whittaker & Wilbur, 1993; Sellen, 1992; 1995; Whittaker & O'Conaill, 1993; 1997). These relationships among affordances, communication behaviours and their effects on core communicative phenomena are shown in Table 2. For example, the Table shows that the visual mode affects multiple communication behaviours, including head nods, gaze and gesture and that one core communication phenomenon affected is turn-taking. Because all theories make predictions about the *differences* between different forms of technology, the table does not show the affordances of linguistic modes. Such information would be redundant, as the linguistic mode is common to all technologies. We now describe three different types of mediated communication theory with respect to these two tables. We

.
.
.

discuss in more detail how different affordances affect behaviours and hence communication process, content and outcome. The remainder of the review will be organized around these three theories: the bandwidth hypothesis, cognitive cueing and social cueing.

| AFFORDANCE TYPE | COMMUNICATION BEHAVIOURS AFFECTED BY AFFORDANCE | CORE COMMUNICATIVE PHENOMENA AFFECTED |
|---|---|---|
| VISUAL MODE | Facial expressions | Attention, understanding, agreement.<br>Conveying affect, attitude. |
| | Head nods | Signal attention, understanding, agreement.<br>Turn-taking. |
| | Gaze | Attention.<br>Turn-taking, reference.<br>Conveying affect, attitude. |
| | Gesture | Attention.<br>Turn-taking, reference. |
| | Visual access to objects in a shared physical environment | Reference and attention. |
| | Physical presence | Availability and initiation of impromptu conversation. |

| | | |
|---|---|---|
| INTERACTIVITY | Feedback via backchannels, completions, interruptions | Attention, understanding, agreement. Turn-taking, reference, repairs. Socio-emotional feedback. |

Table 2 – The effects of different affordances on communication behaviours and processes.

One early theory, the *bandwidth hypothesis* addressed how modes affected communication. The bandwidth hypothesis proposed a direct relationship between the modes that a technology supports and the communication that results from using that technology. The hypothesis posits that, regardless of task, the closer the set of modes supported by a technology approximate to those of face to face communication, the greater the efficiency of the communication using that technology. We should therefore expect a technology that supports both visual and linguistic modes should always outperform one supporting only the linguistic mode. In section 2 we review evidence showing that the bandwidth hypothesis is false.

Subsequent theorizing has been more directly influenced by current psycholinguistic theory. A large number of recent studies have examined the effect of using mediated communication on various types of *cognitive cueing* (Clark & Brennan, 1991, Doherty-Sneddon, Anderson, O'Malley, Langton, Garrod & Bruce, 1997; Isaacs & Tang, 1993; Kraut, Galegher, Fish & Chalfonte, 1992; Kraut, Miller & Siegel, 1996; McCarthy et al., 1991; 1993; O'Conaill et al., 1993; Olson, Olson & Meader, 1995; Sellen, 1992; 1995; Whittaker et al., 1991; 1993; Whittaker, 1995; Whittaker, Terveen, Hill & Cherny, 1998; Whittaker & O'Conaill, 1993; 1997). Section 3 reviews a number of hypotheses about the effects of visual information on communication in cognitive tasks. Following predictions derivable from Table 2, concerning the relations between the visual mode, communication behaviours and core communication phenomena, we examine whether technologies that do not provide visual information disrupt turn-taking (Jackson et al., 2000; O'Conaill et al., 1993; Sellen, 1995; Whittaker, 1995; Whittaker & O'Conaill, 1997) or reference (Bly, 1988, Whittaker et al., 1993), as well as initiating impromptu conversations (Fish, Kraut, Root & Rice, 1992; Tang, Isaacs, & Rua 1994; Tang & Rua, 1994). We also look at how non-interactive technologies disrupt turn-taking (O'Conaill et al., 1993), reference (Oviatt & Cohen, 1991) and

.
.
.

understanding for cognitive tasks (Kraut, Lewis & Swezey, 1982; Oviatt & Cohen, 1991).

A second research area has been into *social cueing* processes in mediated communication. Here the focus is on the role of visual information in supporting the transmission of affective and interpersonal information, as well as the function of interactivity in providing social feedback. Again following predictions derivable from Table 2, we review the evidence for several theories which argue that the absence of visual information changes the expression of affect, and its role in communication (Sproull & Kiesler, 1986; Morley & Stephenson, 1969; 1970; Rutter, 1987; Short et al., 1976). We also examine how the absence of such visual interpersonal information combined with reduced socio-emotional feedback affects higher level social processes such as negotiation (Morley & Stephenson, 1969; 1970), and reaching consensus (Kiesler et al., 1984).

Methods in mediated communication research

Mediated communication research has employed a variety of methods: laboratory studies, field trials, interviews, surveys, and ethnographic techniques. Most of the data reported in this chapter comes from laboratory studies where users are given predefined technologies, tasks and instructions. Often such experiments collect both objective data (e.g. time to complete the task, quality of solution), and subjective data (how useful was technology Y?, how easy was task Z?). Researchers also carry out content analyses of conversations using different mediated communication technologies. Clearly such experiments offer control and the ability to systematically manipulate relevant variables. However other research reported here includes field trials of novel systems where experimental control is sacrificed for data about usage patterns and subjective reactions to technology used for extended periods of time. One critical element of such field trials is that the systems are used for people's everyday work. While such ecological validity is highly valuable, it obviously reduces experimental control over the tasks for which the system is used. Nevertheless, such extended experience with mediated communication systems is vital given recent empirical research showing that many reported differences between mediated communication and face to face communication are vastly reduced after extended periods of technology usage (Walther, 1992; 1994; in press; Kelly, Futoron & McGrath; 1990; Kelly and McGrath, 1985). Other research has employed survey

.
.
.

techniques to examine reported patterns of usage for established technologies such as email and voicemail, e.g. to understand reasons for using these technologies in certain situations (Rice, 1993; Markus, 1987; 1994). A final (small) class of data reported here is ethnographic. Ethnographies employ descriptive, interview and observational techniques to generate fine-grained analyses of communication behaviours, technology usage and reactions to technologies (Nardi, Kuchinsky, Whittaker & Schwarz, 1996; Nardi, Whittaker & Bradner, 2000; Whittaker & Sidner, 1996). The weakness of such techniques concerns their generality, because ethnographic data often investigates only small numbers of users. However ethnographic data differs from other methods in offering detailed qualitative findings about users' techniques and strategies for using various technologies.

We now review evidence for each theoretical approach.

## 2. The Bandwidth Hypothesis

*Hypothesis: The closer the modes supported by a technology correspond to those of face to face communication, the more efficient the communication with that technology. Specifically, adding visual information to speech should improve the efficiency of communication.*

Early research in mediated communication focused on technology modes. It took as its starting point the five human senses (sight, hearing, smell, touch, taste) in face to face communication. Ryan and Craig claimed that different technologies can be characterized in terms of their correspondence with face to face communication, in which *"information can be transmitted and received by any of the five senses"* (pp2-3, cited in Williams, 1977). In mediated settings, sensory input is clearly impoverished. Technologies do not support taste or touch, and visual information is often missing. According to the theory, this should result in reduced efficiency of communication, independently of task. It also follows that technologies supporting multiple senses should be more efficient than those providing access to few senses. We refer to this as the *bandwidth hypothesis.* As we shall see, the work on the bandwidth hypothesis has exclusively concerned the senses of sight and hearing, by investigating the effects of adding visual information to auditory communication. There has been no work into the effects of other sensory modes.

The bandwidth hypothesis was disproved by work examining the effects of various combinations of

communication modes on the process and effectiveness of mediated communication (Chapanis, Ochsman, Parrish & Weeks, 1972, Chapanis, Parrish, Ochsman & Weeks, 1977). This work showed that adding visual information to speech does not necessarily lead to performance improvements. It involved a variety of dyadic, synchronous, communication tasks. Success was measured in terms of quality of solution and time to solution. The set of technologies evaluated was a permutation of face to face communication, remote speech, remote handwriting, remote teletyping, and video/speech. (The term video/speech will be used throughout this review to refer to the combination of video information with speech. We prefer this to the term "video" which can be ambiguous as it is often used to refer exclusively to the visual mode of an audio/visual communication). Communication efficiency was tested for 10 different technology combinations (e.g. speech with remote handwriting, handwriting with remote teletyping, etc.), and for a variety of cognitive tasks that required information exchange and problem solving. On one task, participants had to jointly construct a mechanical object, when one partner was given the unconstructed physical components, and the other, the instructions. In another task, one person had to identify a map location satisfying a number of criteria, e.g. the nearest dentist to a given street address. Again one participant had the map, and the other the relevant selection criteria. Participants were at different physical locations (except of course when the communication condition was face to face communication) and communicated only with the technologies provided.

There were two main conclusions of these multiple studies. First, adding the visual mode to speech to increase "bandwidth" did not necessarily lead to more efficient communication. Face to face communication or video/speech was no more effective than speech only communication for these tasks. These are striking results given that many of the tasks used in the experiments (e.g. finding a location on a map, jointly constructing a complex physical object) would intuitively seem to require access to visual information, and hence to favour mode combinations that transmitted visual data. The result is also of general significance, because it shows that mediated communication is not always different from face to face communication. The second major conclusion was that speech is the critical medium for these types of task. Mode combinations that included speech were always more efficient than those that did not (Chapanis

.
.
.

et al., 1972). If participants used spoken interaction, then the addition of high quality video, text, or writing modes made little difference to task outcome or process. Even face to face communication was no different from speech in task outcome or solution quality. Furthermore, modes that included speech were 2 to 3 times more efficient than non-speech modes.

These results showing little impact of visual information on cognitive problem solving tasks have been replicated in many other laboratory experiments (Reid, 1977; Short et al., 1976; Williams, 1977). For example, Reid (1977) reviewed evidence from 28 different studies showing no effects for the addition of visual information to speech for cognitive tasks. This is not an issue of video quality: even face to face communication is no better than speech only communication for this type of task (Reid, 1977; Williams, 1977). These studies do, however, show some effects for the addition of visual information in tasks in which social cueing is critical, and we review this in Section 4.

Data from naturalistic studies also show that having access to visual information does not always improve communication, when speech is available. A field trial of high quality videophone conducted over several months showed few objective usage differences compared with the telephone (Fish et al., 1992; Fish, Kraut, Root & Rice, 1993). The videophone supplemented speech with a 12 inch television quality image of the remote participants on the user's computer screen. Overall there were few differences between speech and video/speech for either logged usage or user perceptions. Phone and videophone calls had similar durations, and were used for the same set of communication tasks. The researchers also administered a questionnaire asking people to state the tasks for which they felt that different communication 'technologies' (e.g. videophone, telephone, face to face communication) were appropriate. Multidimensional scaling techniques applied to people's answers indicated that videophone is viewed by users as more similar to the telephone than face to face communication. Again this suggests that visual information adds little for this type of task.

In conclusion, there is very little general evidence for the bandwidth hypothesis. A huge amount of data seems to support two complementary and counterintuitive conclusions. The first is that adding a mode does not increase communication performance: specifically visual information generally seems to

.
.
.

contribute little to the outcome of a variety of cognitive tasks, when speech communication is available. The second is the primacy of speech: adding or removing other modes has little effect compared with the addition or removal of speech (Chapanis et al., 1972; 1977).

## 3. Cognitive Cueing

The above results offer strong evidence against the global claims of the bandwidth hypothesis. A more recent set of studies have argued the need to apply face to face communication models more directly to mediated communication (Clark & Brennan, 1991; Daly-Jones, Monk & Watts, 1998; Doherty-Sneddon et al., 1997; Isaacs & Tang, 1993; McCarthy et al., 1991; 1993; O'Conaill et al., 1993; Olson et al., 1995; Tang & Isaacs, 1993; Whittaker, 1995; Whittaker & O'Conaill, 1997; Whittaker et al., 1991; 1993). Consistent with the approach illustrated in Tables 1 and 2, this work claims that instead of focusing on the five senses, we need to identify the contributions of various communication *behaviours* (e.g. speech, gaze, gesture and backchannel feedback) in supporting core communication phenomena, such as turn-taking, reference and understanding. If we can determine how the affordances of various mediated communication technologies affect communication behaviours, we can make more precise predictions about how different types of technologies affect communication. Rather than focusing on general effects of modes, this work argues that we should specify precise mechanisms by which affordances affects behaviours which in turn influence core communication phenomena.

These studies have tested 3 specific hypotheses about different functions of visible information on various cognitive aspects of mediated communication. These concern turn-taking, conversation initiation and the role of a shared environment. They are all directly derivable from the information presented in Table 2. The *turn-taking* hypothesis argues that visible behaviours such as gaze and gesture support turn-taking in face to face communication. Removing these (e.g. in phone conversation) results in impairments in the ability to switch speakers. The *initiation* hypothesis evaluates the importance of visible information about the presence and availability of other conversational participants for impromptu communication. The hypothesis claims that the initiation of impromptu conversations is impaired in technologies (such as the phone) that do not support this type of visual information. The *shared environment* hypothesis (McCarthy

. . .

et al., 1991; 1993; Whittaker et al., 1991; 1993) maintains that a shared physical environment with mutually viewable objects and events contributes to conversational common ground (Clark & Marshall, 1981). The shared environment supports various types of conversational inferences, as well as allowing certain types of reference such as deixis. The absence of the shared environment (e.g. when conversing by phone) will make such conversational inferences harder, reference will be much less efficient, and conversations more explicit.

Other cognitive cueing work concerns interactivity. The *interactivity* hypothesis argues that technologies such as voicemail or email that prevent immediate feedback via backchannels, completions and interruptions, will decrease shared understanding and disrupt various reference processes.

In what follows for each hypothesis, we first characterise relevant behaviours in face to face communication, and describe how they affect turn-taking, availability, reference, and interactivity respectively. We then predict how removing visual information or using non-interactive technologies will affect mediated communication, compared with face to face communication.

Turn-taking

*Hypothesis: technologies that preclude access to visual information about gaze, gesture and head nods will disrupt turn-taking, given the demonstrated role that these behaviours play in managing transitions between speaking and listening in face to face communication.*

There is a great deal of work elucidating the role of gaze and gesture in mediating turn-taking in face to face communication. Speakers and listeners in face to face communication show different patterns of gaze, with listeners spending more time looking at speakers than vice versa (Argyle & Cook, 1976). In addition, gaze can mediate transitions. Speakers tend to look more at listeners as speakers draw to the end of their turn, to await confirmatory information that the listener is ready to continue (Kendon, 1967). Gesture can also serve to co-ordinate turn-taking transitions. The termination of gesture acts as a signal that the speaker is ready to hand over the conversational floor, and is therefore a "turn-yielding" cue (Beattie, 1978, 1981, Duncan, 1972).

The results are mixed for the turn-taking hypothesis. Sellen (1992, 1995) conducted a series of

.
.
.

laboratory studies, in which groups discussed contentious issues and tried to reach consensus. They communicated using speech, video/speech or face to face interaction. As expected, there were differences in turn-taking between face to face communication and speech. More importantly for the hypothesis, however, contrary to expectations, video/speech was no different from speech only communication. For measures such as pausing, overlapping speech and interruption management, there were no differences between the video/speech systems and speech only communication. When compared with face to face communication, both video/speech and speech systems showed reduced ability of listeners to spontaneously take the conversational floor, as measured by number of interruptions. Both mediated communication systems also led speakers to use more formal techniques for relinquishing conversational initiative, such as naming a possible next speaker or using "tag" questions (e.g. "isn't it?", "aren't they?", "couldn't you?", involving an auxiliary verb and question syntax, at the end of an utterance). Consistent with this, O'Conaill et al (1993) and Whittaker & O'Conaill (1993) also found that speakers in naturalistic video-mediated meetings used more formal turn-taking techniques than were observed in face-to-face interaction. They also found that listeners showed reduced ability to spontaneously take the conversational floor. Similar data showing differences in turn-taking between face to face communication and video/speech are reported by Cohen (1982), and Isaacs and Tang (1994).

More recent work has examined the turn-taking hypothesis in multiperson interactions (Jackson et al., 2000). The prediction was that turn-taking problems would be exacerbated in multiparty video/speech interactions compared with dyadic video/speech interactions. This is because the presence of multiple participants should both lead to increased competition for the conversational floor as well as exacerbating the difficulties in tracking gaze and gestural behaviours of all participants. It was therefore hypothesized that reduced video quality induces more turn-taking problems in multiparty than dyadic groups. However the study showed the opposite, with reduced video quality having greater effects on dyadic than four person interactions.

However, subjective data offer more support for the role of visible information in facilitating turn-taking. Sellen (1992, 1995) collected subjective data about turn-taking gathered from questionnaires

.
.
.

addressing subject's impressions of the impact of video on conversational processes. Video/speech was perceived to be better than speech in a number of ways. It was thought to: (a) support interruptions; (b) lead to more natural conversations that are more interactive; (c) increase the ability to listen selectively to particular speakers; (d) allow one to determine whether one is being attended to; and (e) to generally keep track of the conversation. People also believed that they were better able to track the attention of others. Similar qualitative data are reported by Isaacs and Tang (1994), who found that video users felt that they were better able to manage pauses in video than speech only communication. Despite this, Tang and Isaacs (1993) found that high quality video was again not perceived as equivalent to face-to-face interaction: subjective data showed that video was not seen as being as effective in supporting interactivity, selective attention, and the ability to take initiative in the conversation.

Availability in opportunistic conversations

*Hypothesis: technologies that preclude access to visual information about people's presence and availability will reduce the ability to initiate impromptu workplace conversations, given the demonstrated role that visual information about recipient availability plays in opening such conversations. Providing visual availability information should lead to fewer failed attempts to initiate impromptu conversations.*

The second cognitive cueing hypothesis concerns the role of visible behaviours in initiating impromptu communications. One can infer the presence of another person (and hence their potential availability for communication) if they are visible. Information about the proximity, current activities and movements of other people affect various aspects of impromptu conversations, (e.g. their initiation), and termination as well as how interruptions are handled (Heath & Luff, 1991; Kendon & Ferber, 1973; Kraut, Fish, Root & Chalfonte, 1993; Tang & Rua, 1994, Tang et al., 1994; Whittaker, Frohlich & Daly-Jones, 1994; Isaacs, Whittaker, Frohlich & O'Conaill, 1997). Workplace interactions are generally unplanned (Kraut et al., 1993; Sproull, 1983; Whittaker et al., 1994), and visible information provides mechanisms for initiating such impromptu conversations. Firstly, sightings of others can lead one to fall into spontaneous conversation, (e.g. in public areas such as hallways) (Kendon & Ferber, 1973). In addition, seeing a colleague may remind one of an issue that needs to be discussed (Kraut et al., 1993, Whittaker et al., 1994).

.
.
.

Visible information is also helpful in determining whether colleagues might be receptive to an unplanned conversation, offering vital clues as to how available or interruptible they are. Clearly all these types of visible availability information are absent from phone only conversations. One striking statistic is that 60% of business phone calls fail to reach their intended recipient (Rice & Shook, 1990). Calls are either unanswered, or transferred to voicemail or a secretary. This failure of the phone has been attributed to the absence of availability information, in contrast to face to face settings which allow participants to time unplanned phone calls for when the recipient is present and receptive to conversation.

Three separate classes of video application have been built to provide availability information to facilitate impromptu interactions between remote coworkers: (a) *glance* which enables a user to briefly "look into" the office of a co-worker to assess their availability; (b*) open links* in which persistent video/speech channels are maintained between two separate physical locations simulating a shared physical office, allowing constant availability information; and (c) *awareness* applications in which video images of coworkers' offices are periodically sampled, so that "snapshots" of their office can show their recent movements and availability. The difference between awareness and glance is that awareness information is not continuous: it may be a single frame updated periodically.

Fish et al., (1992, 1993) tested the use of different types of *glance*, and their differential success in promoting impromptu interactions. Brief glances at a user selected recipient were the most frequently chosen type: 81% of user initiated interactions were of this type, but only 54% of these initiated conversation. This rate is no better than with phone only interaction. All other types of glances were much less frequent had even lower likelihoods of resulting in conversations. One type of glance was intended to simulate chance meetings such as "bumping into" another person in a hallway. This showed very high failure rates, with 97% being terminated immediately without conversation. The relationship between glances and impromptu communication was also explored by Tang et al. (1994) for a system operating across multiple sites in a local area. Participants could first "look into" the office of a remote coworker, with the option of converting this into an extended conversation. Altogether, only 25% of glances were converted into conversations. Again, this is worse than connection rates using only the phone.

.
.
.

Permanently open video links support the ability to "waylay" a potential recipient who is out of their office. By monitoring the link one can see when they return, ensuring that a vital communication takes place (Bly, Harrison & Irwin, 1993; Fish et al., 1992; Heath & Luff, 1991; Mantei, Baecker, Sellen, Buxton, Milligan & Wellman, 1991). There are also claims that working with a video link may be less intrusive than sharing a real office, but still offer many of the same benefits, in terms of access to other participants (Heath & Luff, 1991). Evaluation data suggest, however, that open links may be the exception rather than the rule. Fish et al (1992) report that only 5% of connections lasted more than 30 minutes, and Tang and Rua (1994) report only 2% lasted more than 30 minutes. Furthermore, in both cases, these data may overestimate the frequency of open links, because they include extended continuous interactions as well as intermittent conversations over open links. Thus, both sets of usage data suggest that brief interactions, rather than open links are the main uses of the system, again suggesting that availability data may not be enormously valuable.

Open links can also be constructed between public areas (Abel, 1990, Bly et al., 1993, Fish, Kraut & Chalfonte, 1990; Root, 1988). Cameras were installed in common areas, transmitting images of people who were in equivalent public areas at remote sites. This was intended to promote impromptu conversations of the type that can occur when people meet in public areas of the same site, (e.g. coffee room or water cooler conversations). These studies report the frequent use of open links for social greetings between remote sites, with 70% of usage being of this type (Abel, 1990, Bly et al., 1993). However, the use of the open link was mainly limited to these brief social exchanges, and the link was seen by users as being ineffective in supporting work (Fish et al., 1993). Furthermore, sightings over a videolink were again less likely to convert to extended conversation than face-to-face sightings (Fish et al., 1990).

Overall, these results offer little evidence for the availability hypothesis. Nevertheless there are methodological problems with the studies reported. These negative results may have occurred because of the weakness of particular implementations. Video quality was often low and it is well-documented that conversations over low quality video are disrupted compared with face to face communication (Cohen, 1982; Daly-Jones et al., 1998; O'Conaill et al., 1993). People may therefore be unwilling to initiate

.
.
.

conversations when these will be held over low quality video.

Shared environment and shared attention

*Hypothesis: technologies that preclude access to visual information about gaze and gesture in a shared physical environment will disrupt reference, given the demonstrated role that these behaviours play in supporting reference in face to face communication. Reference should be less efficient, taking more turns with speech only communication and deixis should be much reduced. Communication using a shared environment should also be more efficient and more implicit for tasks requiring physical manipulation, as the shared environment provides a direct method to jointly perceive changes to the environment which therefore do not have to be referred to directly.*

In face to face communication, the fact that participants have access to a *shared physical environment* means that other types of visible information are available, such as shared information about physical objects and events. Sharing the same physical environment enables people to coordinate conversational content, by making inferences about the set of objects and events that others in the same environment are likely to know about and want to talk about (Argyle & Graham, 1977; Clark & Marshall, 1981; Cooper, 1974; Kahneman, 1973; Whittaker et al., 1991; 1993; 1994). Information about the visible environment often interacts in important ways with verbal and visible behaviours, (e.g. when participants gesture at, orient towards, and manipulate aspects of their environment). A crucial aspect of conversational content co-ordination is the ability to achieve joint reference, and gaze can play a role in reference to external objects or events. People are very good at determining where others are looking (Bruce, 1995, Watt, 1995). Gaze serves to co-ordinate the joint attention of conversational participants. When speakers and listeners achieve joint attention to an object, event or location, speakers can refer to it by pointing or deixis. Even when they do not resort to pointing gestures, joint attention to an object or event also allows participants greater flexibility in how they verbally refer to it (Clark & Marshall, 1981). Joint attention also contributes to inferences about common ground. If both participants have observed a change to an object or event, they can assume that such changes are part of the conversational common ground. They therefore do not have to be mentioned explicitly.

.
.
.

A number of studies have examined the role of a shared visible environment in remote design tasks requiring depiction and spatial reasoning. Bly (1988) observed dialogues in 3 different technology configurations: face to face communication, speech only and speech/video. In all cases, participants were provided with paper on which they could write or draw, and in the speech/video condition the video camera was directed at the writing/drawing surface, so that it could be seen by the remote participant. In the face to face and speech/video conditions, participants could therefore observe the other person's writing, drawing and gesturing behaviours, whereas in the speech condition they could not. Bly observed the relative frequencies of gesturing versus writing and drawing in the 3 conditions. Levels of gesturing were high when there was a shared visible environment: 61% in face to face and 31% in speech/video, but interestingly, gesturing still occurred 11% of the time in the speech only condition. Deixis was frequently used to reinforce talk: people would underline or circle objects that they were currently discussing, even when they did not refer to these with linguistically deictic forms such as "this" and "that".

A study by Whittaker et al., (1993) compared communication effectiveness when people interacted using speech, with and without an electronic shared workspace for 3 tasks, design layout, collaborative text editing and brainstorming. The shared workspace enabled people to share visual material such as documents or designs as well as to type, draw, write and annotate these. In the design layout task participants had to configure a room by arranging "pieces of furniture" of different shapes and sizes given a room plan of specified dimensions. Pieces were given different scores and participants had to maximize their score joint score by inserting high value pieces. Participants obtained higher scores faster when they had the workspace. Conversational processes also differed between the two conditions. Participants were able to depict and identify visual objects more easily with the workspace, because it allowed the straightforward expression of spatial relationships and locations, and this was often achieved by gesture. Efficiency of reference was demonstrated by the finding that participants with the workspace took fewer turns to identify a given piece and fewer turns overall. This was partly because participants were able to refer to pieces deictically. Similar effects were observed in the collaborative editing task. Here the workspace facilitated the straightforward gestural depiction of spatial relationships for text edits that would

.
.
.

have been laborious to describe, such as where a sentence should be relocated within a paragraph, or where a paragraph should be moved. In contrast, for the simple brainstorming task there were few benefits for the workspace, because of the absence of a spatial component to the task. Olson et al., (1995) reported similar results for a complex visual design task. They compared speech, speech/video and face to face interaction for visual design tasks. Face to face groups generated higher quality designs than speech only groups, but only when they had access to a shared workspace, allowing participants to jointly view, annotate and gesture at, shared designs. There were also differences in conversation processes. Speech only groups spent more time in stating and clarifying issues, again suggesting the utility of the workspace in supporting common ground.

Kraut et al. (1996) also investigated the effects of a shared visual environment provided by video, on the performance of the physical task of repairing a bicycle. Novices had to make the repairs to the bicycle while connected to a repair expert by one of three technology combinations. These were (a) speech; (b) speech and high quality video; (c) one way audio and high quality video. There were no overall differences in quality of solution in the technology conditions, but there were differences in the *process* of advice giving. In the two video conditions, help-giving was more proactive: experts did not have to be prompted by novices to receive help. In addition, novices were much more implicit in their descriptions of the state of the repair. These observations can be explained in terms of common ground. In the video condition, the experts could perceive when the novice was experiencing set up problems without having to be told this. Also novices were able to assume that experts could see the current state of the system without having to describe it in detail. In both cases, having shared common ground allows participants to be more implicit about what needs to be said.

These findings are also confirmed by ethnographic research looking at the use of video images to share information about patients' brains and spines during neurosurgery (Nardi, Schwarz, Kuchinsky, Leichner, Whittaker & Sclabassi., 1993; Nardi et al., 1996; Whittaker, 1995, Whittaker & O'Conaill, 1997). We compared communication in surgical practice with and without the provision of this visual information. In one set up, the team assisting the surgeon had no visual information about the actions that the surgeon

.
.
.

was currently carrying out on the patient. In the other, we provided all team members with broadcast images of the surgeons' actions on the patient's brain or spine produced by a magnifying microscope. We observed two different types of communicative use of the video image. First, the dynamic image of the surgeon's actions on the patient allowed detailed co-ordination of interleaved physical action between the assisting nurse and the operating surgeon. By monitoring the surgeon's current actions, via a shared video image viewed through the microscope, the assisting nurse could anticipate the surgeon's requirements and provide the correct surgical instrument, without it being directly requested. A second communicative function of the video image was that it served to disambiguate other types of surgical data that were supplied to remote consultants, such as neurophysiological monitoring data. The interpretation of these neurophysiological data depends critically on precise information about the physical actions that the surgeon is currently executing, such as the exact placement of a surgical clamp. Without the video image depicting these actions, the remote consultant had to rely on telephone reports from those who were present in the operating theatre, and the inadequacy of the descriptions meant that the consultant often had to resort to physically visiting the operating theatre to observe the actions directly. In both cases the video provided situational information that was previously unavailable to certain participants. They no longer had to ask for this information directly, allowing communication to be more implicit and hence more efficient.

Interactivity

*Hypothesis: technologies that preclude access to listener feedback in the form of backchannels, completions and interruptions will disrupt reference, repairs and understanding given the demonstrated role that these feedback behaviours play in face to face interaction. Technologies that fail to support interactivity will show less incremental reference, more redundancy, longer turns, reduced interruptions and reduced speaker switching.*

The importance of interactivity in face to face communication is well documented. Listeners offer speakers regular feedback as to whether conversation is on track, by giving verbal evidence in the form of backchannels ("mm", "uhu") and visual evidence in the form of head nods and attention (Clark, 1992; 1996; Clark & Wilkes-Gibbs, 1986; Duncan, 1972; Kendon 1967, Krauss, Bricker, McMahon & Garlock,

1977; Kraut et al., 1982, Schegloff, 1982; Walker & Whittaker, 1990; Whittaker & Stenton, 1988, Yngve, 1970). Feedback is often given concurrently with the speaker's utterance to indicate understanding, and in some cases acceptance, of what the speaker has said (Clark & Shaefer, 1989). In addition, the fact that listeners can interrupt to clarify or question the moment that they experience breakdown of understanding has been shown to be an important mechanism for achieving shared understanding (Clark & Wilkes-Gibbs, 1986, Clark & Shaefer, 1989, Oviatt & Cohen, 1991, Schegloff, 1982). A different and stronger sense in which conversation is interactive is that all participants have the chance to incrementally contribute to, negotiate and modify the content of what is being said. Thus listeners will sometimes interrupt to clarify what a speaker meant, dispute an assertion, ask a question, or complete an utterance for the speaker. Speakers also offer frequent opportunities for listeners to make this type of contribution.

Despite the crucial role of interactivity in face to face communication, there has not been a huge amount of research into interactivity in mediated communication. The lack of research is surprising because pervasive technologies such as email and voicemail do not support interactivity and it would seem to be important to document how this affects communication using those technologies. One of the few attempts to directly test the effect of interactivity is an experiment by Oviatt & Cohen (1991). This directly compared communication processes and outcome using interactive and non-interactive versions of the same linguistic communication mode. Experts provided instructions to novices about how to assemble a toy water pump, either interactively by telephone or non-interactively by recording an audio tape (analogous to leaving a voicemail message). In the audiotape condition, novices listened to the recorded instructions using a tape recorder which they could stop and start as they wished, as they carried out the expert's instructions. In the interactive condition, instructional exchanges occurred in real-time, allowing incremental feedback. The results showed that interactive communication was more efficient. Interactive pairs were much faster to complete their tasks. Analysis of communication processes also showed that in the interactive condition there was moment by moment coordination of communication. Definite reference was often incremental (Clark & Wilkes-Gibbs, 1986), with the novice confirming at each stage of the reference process that they were following the description. This incremental feedback allowed

.
.
.

misunderstandings to be quickly redressed. In contrast in the non-interactive condition, there were large inefficiencies of communication. Experts tended to generate highly redundant instructions: reference descriptions were often elaborated, with repeated descriptions being applied to the same object. Experts also tended to generate more structural and summary remarks in the non-interactive condition. Oviatt & Cohen argued that this over-elaboration and structural marking was a compensatory response to the absence of incremental feedback. Without feedback, experts in the non-interactive condition could not establish whether they had been correctly understood. In consequence, they over-elaborated instructions to block potential misunderstandings. Similar results of over-elaboration are reported by Krauss and Weinheimer (1966) for a comparison of interactive versus non-interactive speech.

Another test of interactivity is to investigate the effects of preventing conversational feedback in face to face, video/speech or speech communication. Several laboratory studies have shown that the absence of persistent feedback also leads to over-elaborate and highly redundant messages (Krauss et al., 1977), as well as decreased mutual understanding (Kraut et al., 1982). Another source of evidence for the importance of interactivity comes from studies of delayed, or disrupted, speech communication. Disruptions to interactivity can occur either because the speech channel is unidirectional (preventing concurrent feedback, completions and interruptions) or because of transmission lags (resulting in feedback or interruptions being delayed and arriving at inappropriate times). Interactive phenomena such as backchannels, repairs, and interruptions characteristically require very precise timing and minor disruptions to the transmission characteristics can seriously affect these (Jaffe & Feldstein, 1970; Norwhine & Murphy, 1938).

Krauss and Bricker (1967) investigated the effects of pure transmission delays and half-duplex (unidirectional) speech in phone conversations. In the half duplex condition, only one person could speak at a time and listeners could not interrupt until the previous speaker had been silent for delays between 250ms. and 1.8s. Both manipulations were found to have disruptive effects, leading to speaker overelaboration. These were more serious in the half-duplex case: lack of bidirectional access to the channel was disruptive at intervals as short as 250ms. Similar data showing the effects of delayed feedback are reported in studies of videoconferencing. Cohen (1982) investigated the effects of a 705ms. lag on

.
.
.

various conversational measures. (This duration of lag was characteristic of videoconferencing systems in 1982). The lag led to longer conversational turns, and reduced overlapping speech when compared with face to face communication. The absence of overlaps was interpreted as compromising key aspects of interactivity that are necessary for concurrent feedback, turn completions and incremental reference. Whittaker and O'Conaill also investigated these underlying processes in a similar, naturalistic study of videoconferencing, comparing lagged duplex videoconferencing, with zero lag, full duplex videoconferencing and face to face communication. Half duplex speech and lags of between 410 and 720ms. led to reduced backchannelling, fewer interruptions, fewer completions and fewer overlaps (O'Conaill et al., 1993; Whittaker & O'Conaill, 1993). Listeners found it harder to time incremental feedback or to interject in a manner that was not disruptive to the speakers. They would therefore wait until the current utterance was complete before making their contribution. Lack of feedback led to longer speaker turns, because speakers tended to overelaborate.

Finally, comparisons of spoken and typed interactions also demonstrate the effects of reduced interactivity. Interactive behaviours require very precise timing and the lack of fluency of typed communication should disrupt their execution. Oviatt and Cohen (1991) compared differences between spoken and typed conversation for the toy water pump construction task described above, involving experts and novices. Typist experts generated instructions that were more verbose and more complex. Typist novices produced many fewer backchannels and clarifications. This lack of interactivity led to more elaborations and redundancy in typed than speech instructions, leading pairs of typists to take 3.6 times as long to complete their tasks.

Summary of Cognitive Cueing Hypotheses

Evidence for the cognitive cueing hypothesis is mixed. While there is strong evidence for the effects of *interactivity* and *shared environment*, the results for *turn-taking* and *availability* are much less convincing. There are two potential reasons for the negative findings; (a) poor implementations of video systems; and (b) misanalysis of the role of visual behaviour in communication. First, the failure to find turn-taking and availability effects may result from flawed implementations, because current video

.
.
.

implementations may not capture important aspects of face to face visual behaviour. For example, mutual gaze is impossible, and proxemics and gesture are distorted (Heath & Luff, 1991; O'Conaill et al., 1993; Sellen 1992; 1995, Whittaker, 1995).

Another, more compelling, argument for the negative results concerns the role of visual information in communication. Most of the experimental research on cognitive cueing (e.g. the turn-taking and availability hypotheses) concerns the role of visual information *about other conversational participants*. Yet the results suggest that for cognitive tasks, it is more important to show a *shared environment* rather than to depict other participants. Why might this be the case? One possibility is that participant information is unimportant for this type of task, because participants spend relatively small amounts of time gazing at others, in the presence of relevant visible objects. Argyle (1990) reported that it is highly unusual for listeners to spend more than 50% of their time looking at others, with other studies reporting much lower frequencies (Anderson, Bard, Sotillo, Doherty-Sneddon, & Newlands, 1997). Gaze at others falls to 3-7% of conversational time when there are interesting objects present (Argyle & Graham, 1977). Mutual gaze (when both participants are simultaneously looking at each other) is even lower, falling to below 5% (Anderson et al., 1997).

Gaver, Sellen, Heath and Luff (1993) tested what video images people select when they are engaged in a complex physical design task. Users could choose between an image of the other participant, and various views of the object under study. Participants chose views of their co-participant only 11% of the time. Mutual gaze, where both participants were simultaneously viewing each other, occurred only 2% of the time. Instead, people were much more likely to choose an image of the object, spending 49% of their time with the object views. This supports the view that for cognitive tasks, information about the gaze and gestures of others is less important than information about the shared physical context. Given that participants spend so little time looking at others, they have restricted opportunities for eliciting visual information about them. This may explain the lack of evidence for the *turn-taking* and *availability* hypotheses. It also suggests that when building video applications for cognitive tasks, it might be more profitable to provide information about shared objects rather than people (Anderson, Smallwood,

.
.
.

MacDonald, Mullin, Fleming, & O'Malley, 2000; Nardi et al., 1993; 1996; Whittaker, 1995; Whittaker & O'Conaill, 1997).

## 4. Social Cueing

The previous section focused on cognitive cueing in mediated communication. However, communication is not restricted to the exchange of propositional information, and one important aspect of communication concerns the interpersonal and social attitudes of other participants, including their feelings, emotions and attitudes. As with conversational intentions, participants generally do not make this information verbally explicit, so that it often has to be inferred. Exchanging this type of information can have significant effects on the content and process of communication. The social cueing hypothesis concerns visible behaviours such as facial expressions and gaze, which convey affective information, and interactive behaviours providing socio-emotional feedback. The hypothesis states that technologies that fail to support interpersonal information and feedback will alter both the emotional content of communication, as well as higher level social processes that mediate negotiation.

There have been two major theories proposed for social cueing. *Social presence* theory (Short et al., 1976) argues that technologies differ in the extent to which they present a sense of other communication participants' goals attitudes and motives. Because face to face communication and video/speech provide rich visual interpersonal information afforded by gaze and facial expressions, they are high on the scale of "social presence", with text at the opposite end of this scale. According to social presence theory, using a technology that fails to communicate social presence will change the content and outcome of communication for tasks that require access to interpersonal information. Another focus has been the absence of *social cues* in mediated communication (Kiesler et al., 1984, Culnan & Markus 1987, and Rutter 1987). This work argues that the lack of visual interpersonal information in mediated communication, combined with attenuated socio-emotional feedback transforms emotional content and also higher level social processes such as negotiation. Although most of the work on social cueing has investigated remote synchronous communication (such as chat), some of the research that we describe here concerns asynchronous settings. For conceptual clarity we describe all this work together.

.
.
.

There are three main *social cueing* hypotheses. These concern (a) communication *content*, (b) *negotiation and deadlock,* and (c) *social processes* such as participation patterns and acceptance. The hypotheses are derivable from Table 2. The *content* hypothesis addresses the role of interpersonal information about participants' affective and attitudinal state, provided in face to face communication by visible behaviours, such as gaze and facial expression, along with socio-emotional feedback provided by backchannels. Technologies that limit access to these behaviours will lead to conversations with different emotional content than face to face communication. The *negotiation and deadlock* hypothesis predicts that technologies that reduce information about participants' attitudes and affect cause participants to overlook other's perspectives. This will change the outcome of tasks such as negotiation that require extensive perspective taking. In extreme cases, the inability to take another's perspective will lead to deadlock. Finally, technologies that limit access to interpersonal information and social feedback are hypothesized to change high level social processes such as conversational *participation pattern, and acceptance* of others' conversational contributions.

Content differences in mediated communication

*Hypothesis: The content hypothesis concerns the role of interpersonal information provided in face to face communication by visible behaviours, such as gaze and facial expression, along with interactive feedback. Technologies that limit access to these behaviours will lead to communication with different emotional content than face to face communication.*

Visible behaviours such as gaze, gesture, and facial expressions contribute to the communication of interpersonal information. Gaze is an important indicator of interpersonal attitude or affect. Gaze patterns normally show specific distributions with few gazes lasting more than a second. As a consequence, any deviation from this distribution is associated with an unusual form of interaction. Speakers tend to gaze at a listener's face more when they are being more persuasive, deceptive, ingratiating or assertive (Kleinke, 1986), possibly because under these circumstances it is crucial for speakers to monitor the effects of their speech closely. People also tend to look more at conversants whom they like (Exline & Winters, 1965). In addition, people tend to evaluate others by their patterns of gaze: people who look at their interlocutor only

. . .

a small part of the time are judged as "defensive" or "evasive", whereas those who look a lot of the time are "friendly" "mature" and "sincere" (Kleck & Nuessle, 1968).

Facial expressions are also highly informative about the affective state of conversational participants and their current interpersonal attitudes. The face is highly visible and conversational participants can interpret a great deal from the faces of others. Ekman and colleagues (Ekman, 1982, Ekman & Friesen, 1975) have shown that people across a number of cultures are able to recognise seven distinct facial expressions from posed photographs (happiness, sadness, surprise, anger, disgust, fear, and interest). Affective expressions can be important for both speakers and listeners. They allow listeners to infer speakers' emotional stance to their utterance, while speakers can determine their audience's reaction to what is being said. The primacy of non-verbal affective information (whether it is conveyed by facial expressions or gaze) is demonstrated by studies showing that when this visual information is in conflict with verbal information, people tend to trust visual information (Short et al., 1976).

Consistent with the hypothesis, there is evidence that the *content* of communication is different when visual information is present for negotiation tasks. Stephenson, Ayling and Rutter (1976) compared two person debates about union management relationships, in speech only, and face to face communication settings. They found that speech discussions were more impersonal and task-oriented. Depersonalisation was manifested by reduced praise for one's opponent, more blame for the opponent, fewer self-references and more overall disagreement. Speech discussions also produced more purely informational interactions containing no reference to other conversational participants. Subjective data also support the content hypothesis. Participants believe that technologies providing visual interpersonal information such as video/speech and face-to-face are better than the telephone for tasks requiring access to interpersonal information, such as getting to know other people, or person perception tasks. In addition, groups conversing using video/speech tend to like each other more and to regard other participants as more intelligent than those using only speech (Reid, 1977, Short et al., 1976, Williams, 1977). Similarly, Rutter (1987) showed that speech discussions are perceived to be more socially distant than face to face discussions.

.
.
.

A number of similar studies have shown that textual communication is more impersonal and task oriented than face to face communication (Culnan & Markus, 1987; Hiltz, Johnson & Turoff, 1986; Hiltz & Turoff, 1978, Rice, 1984). Textual communication obviously does not provide access to visual interpersonal information. These studies compared the content of chat, and email with face to face communication. Consistent with the hypothesis, they showed that people using textual communication are more concerned with task constraints and matters of fact, rather than the feelings and motives of their conversational partners. They also found lower proportions of socio-emotional messages showing agreement. However, these studies argued that the absence of interactive feedback rather than the absence of visual interpersonal information leads mediated communication to show a prevailing task orientation (Hiltz, 1975; Hiltz,, Johnson and Turoff, 1986; Hiltz, Turoff & Johnson, 1989).

A different, and more controversial set of arguments about content in mediated communication concerns "flaming": defined as inappropriate and strong displays of emotion. Several early studies (Siegel, Dubrovsky, Kiesler & McGuire, 1986; Kiesler, Zubrow, Moses & Geller, 1985; Sproull & Kiesler, 1986) argued that such behaviours are more prevalent in mediated communication because the social processes such as feedback or access to interpersonal information that normally moderate these displays are attenuated. Inappropriate displays of affect may also be self-reinforcing: an initial display of inappropriate affect may engender equally strong responses, leading to extended interactions with highly emotional exchanges ("flame wars"). Several early studies report that such behaviours are much more frequent in mediated communication than face to face communication (Kiesler et al., 1985, Sproull & Kiesler, 1986). Sproull and Kiesler (1991) report 102 flaming remarks in 24 electronic discussions compared with only 12 such remarks for the same number of face to face discussions. Chesebro (1985) studied the most recent 10 messages from 14 public computer bulletin boards and found that 32% were interpersonal in nature. Meyers (1985) analysed two bulletin boards and found that jokes, insults, sexual topics, games, stories and personal information constituted 29% and 39% of the content of the systems, respectively. However other more recent studies do not find inflated levels of socio-emotional content in email and chat compared with face to face interaction (Hiltz, Johnson & Agle, 1978; Lea & Spears, 1991; Rice & Love, 1987).

.
.
.

Furthermore in some of these studies there was less flaming in mediated communication than in face to face communication. For example, Hiltz & Turoff (1978) found an average of 14% socio-emotional content in 8 mediated communication groups compared with 33% in 3 face to face communication groups, where the groups were trying to reach a joint decision.

Negotiation and deadlock

*Hypothesis: The negotiation hypothesis predicts that technologies that reduce access to interpersonal information about participants' attitudes and affect make perspective taking difficult, and this combines with the reduction in social feedback to make it harder for participants to negotiate effectively and achieve consensus.*

There is good evidence for the claim that removing visual information leads to more task oriented behaviours in mediated communication. Using video/speech as opposed to speech changes the outcome of communication tasks that require access to interpersonal information, such as negotiation, bargaining, and conflict resolution. Short (1971; 1972) studied negotiation behaviours in three technology conditions: face to face, video/speech and speech only. Participants engaged in simulated role-playing negotiations where they represented either unions or management in an industrial dispute. Each participant was assigned a role and given a detailed description of the dispute. Unknown to the participants, the descriptions were worded so that one side always had objectively the stronger case. These studies demonstrated, counterintuitively, that there were more settlements in favour of the side with the stronger case in the speech condition. There was also a marked tendency for speech conversations to end in deadlock, whereas face to face participants were more likely to compromise. Similar data are reported by Morley and Stephenson (1969; 1970) who compared the outcome of negotiation tasks in face to face, video/speech and speech only conditions. They found differences between the visual and non-visual conditions. Speech only communication demonstrates greater task focus, whereas video/speech and face to face participants are more likely to compromise. The results can be interpreted in the following way: the reduction of interpersonal information in the speech condition causes participants to behave more impersonally. They attempt to pursue task goals, in order to win the negotiation, a situation that sometimes produces deadlock. When deadlock does not occur, the side

.
.
.

with the factually stronger case tends to succeed. In contrast, in face to face communication, the greater availability of interpersonal information leads participants to become less task focused. Here, the merits of the case are less prominent and compromise is more likely to occur (Reid, 1977; Short et al., 1976; Williams, 1977).

Other research investigates the processes for arriving at consensus in mediated communication. Decision making in face to face settings generally shows a gradual convergence with successive speakers being more likely to agree than disagree with the proposals of prior conversants. In mediated communication, in contrast, the absence of visible interpersonal information and social feedback make rapid convergence harder to achieve. Hiltz et al., (1986) showed that for judgment tasks, mediated communication groups using textual conferencing were less likely than face to face communication groups to reach consensus. They also had more difficulty in deciding among alternatives. Other related research shows that coalitions are easier to form when visual information is available whether this is video/speech or face to face communication (Reid, 1977). This is again consistent with the view that visual information provides access to the feelings and motives of others. However another result from the Hiltz et al., (1986) study suggests that these findings may be task-specific. Hiltz et al. found that consensus could be achieved in mediated communication when intellective tasks were used.

Participation and acceptance

*Hypothesis: Technologies that limit access to interpersonal information and social feedback compromise social processes. This leads to equalized conversational participation patterns, and reduced acceptance of others' conversational contributions.*

Multiperson face to face interactions have strong participant inequality. Sproull and Kiesler (1991) show that in small group face to face discussions, half the participants contribute only 10-20% of talk. Furthermore, the amount that people talk is strongly correlated with their perceived status within the group. For example, in face to face groups, managers talk more than subordinates and men more than women. This holds true even when the higher status (and generally more vocal) members are not more expert on the topic under consideration. In contrast, mediated communication groups communicating real-time using text

.
.
.

show much greater equality of participation (Dubrovsky, Kiesler and Sethna, 1991; Hiltz et al., 1986; Hiltz et al., 1989; Kiesler et al., 1985; McGuire et al., 1987; Siegel, et al., 1986; Weisband, 1992). Similarly, subjective data from interview studies of email show equalization of contributions. Email is reported to be more 'democratic', because minorities and lower status individuals feel they are more able to contribute (Sproull & Kiesler, 1991). Email also shows a greater diversity of perspectives (Sproull & Kiesler, 1991). These studies argue that mediated communication attentuates visible social status cues, because personal identity is often restricted to an email address or login name. Greater participant equality partly results from this reduction in visible status cues.

Other research has shown differences in *patterns of acceptance* of others' contributions in mediated communication. Several studies show that consensus is harder to achieve in mediated communication (Hiltz et al., 1978; 1986; 1989; Olaniran, 1994; Siegel et al., 1986; Straus & McGrath, 1994, Rice, 1984). The imposition of status is argued to be one mechanism for ruling out competing alternatives when there is dissent. Two studies support this view. Dubrovsky et al. (1991) documented how status effects are reduced in mediated compared with face to face communication. In mediated communication, high status members are less likely to be the first participant to propose a decision, and, if made, their proposals are not as likely to be accepted as in face to face communication. Wiesband (1992) reports similar conclusions in a study of consensus formation comparing chat and face to face conversations.

Summary of Social Cueing

The results for the specific social cueing hypothesis are again mixed. Although there is good support for the effects of mediation on *negotiation, participation* and *acceptance,* support for the *content* hypothesis is less clear. Although mediated communication often has different *content* from face to face communication, we need to refine this hypothesis to avoid apparent contradictions. In some cases, we are led to expect mediated communication to have *increased* emotional content and interpersonal expression (e.g. flaming), but in other cases it is predicted to show *reduced* interpersonal focus, leading to task-centric behaviours. How can we explain these apparent contradictions? One possibility is that there are different types of emotional effects, and that these arise from different underlying mechanisms. One final

.
.
.

observation about this work is that the data seem to be more consistent with a *task-specific* account, rather than the view that social processes are *generally* different in mediated communication. There are few reports of changed emotional content or changes in *negotiation*, *participation* and *acceptance*, when the tasks are intellectual in nature (Hiltz et al. 1986, 1989, Reid 1977, Short et al. 1976). This is consistent with the earlier research on cognitive cueing reviewed in Section 3.

Recent work on *social throughput* has suggested an alternative mechanism to explain different content effects in mediated communication. It is argued that these arise from a failure of social throughput, because typed communication is used. The argument is that content effects arise from the lack of fluency and slow speed of typing, rather than the absence of visual interpersonal information. Daly (1993) and Walther (1992, 1994, 2002) claim that in textual communication the effort and slow pace of typing cause participants to be more terse, leading them to be perceived as less polite. This, in turn, engenders flaming or dissent. Support for this view comes from studies that give participants who are communicating by typing unrestricted time to complete tasks. Unrestricted time should allow participants to overcome lack of fluency of typing allowing focus on emotional expression. As we shall see, several studies show that, given more time, mediated communication participants adjust their typing to be less terse and more socio-emotional.

Social throughput also has implications for other *social cueing* hypotheses. According to some theories (Rutter, 1987), changes to negotiation processes, and participation levels follow directly from the impersonal content of mediated communication. But if changes to emotional content are the result of textual media failing to provide sufficient social throughput, then it follows that *negotiation, participation and acceptance* behaviours should also change when participants are given more time to complete tasks. This turns out to be the case. Studies show that given unrestricted time, outcome differences between chat and face to face communication are much reduced and that participant inequality reasserts itself in mediated communication (Hiltz et al., 1989; McLeod, 1992). The effects of social throughput are also found in naturalistic studies: both Reid et al. (1996) and Steinfield (1986), show that socio-emotional remarks decrease in mediated communication when groups are under deadline pressure. Again the slow rate of typing leads participants under time pressure to abandon socio-emotional discourse to focus on the

.
.
.

task at hand. Together these data argue that for applications such as chat or email, rather than following directly from the absence of interpersonal visual information, social cueing effects may result from problems of social throughput (Walther, 2002). Nevertheless, social throughput may not be able to explain some early work (Stephenson et al., 1976; Morley & Stephenson, 1969; Reid, 1977) comparing speech and face to face communication. It is harder to explain how these differences could be explained by social throughput, because both media conditions used speech, which is highly fluent. These early studies nevertheless showed differences in content and task outcome between face to face and speech only communication. This suggests there may be two separate types of social cueing effects, arising from lack of social throughput on the one hand, and lack of visual interpersonal information on the other. We need more research to address this.

Taken together, the social cueing results indicate that we need better defined models of how social cueing occurs and experiments that partial out the effects of different mechanisms before we can make systematic statements about how these processes operate. More specific accounts and data might lead us to be able to tease apart some of the differences between social throughput and visual information effects.

## 5. Conclusions

We now present some general conclusions, and suggest future theoretical and empirical work that needs to be done on mediated communication.

One of our initial orienting questions concerned whether face to face and mediated communication are different. Our data show that this depends on the task. For cognitive tasks, using interactive technologies there are generally few differences between face to face communication and mediated communication (Chapanis et al., 1972; 1977). However mediation seems to have larger effects for tasks requiring access to interpersonal information (Short et al., 1976; Kiesler et al., 1984; Morley & Stephenson, 1969), or those requiring complex joint physical manipulation (Bly, 1988; Kraut et al., 1996; Whittaker et al., 1993).

This result is important for theories of mediated communication. How can we explain these effects? The finding that mediated and face to face communication are equivalent under certain conditions is theoretically important, because it shows mediation *per se* does not change communication. Rather, the

.
.
.

results indicate that we need to explain these observations in terms of the underlying affordances of various technologies, and how these affordances affect a technology's ability to support various communication behaviours. The data indicate that these behaviours in turn have effects on core communication processes. Similarly, it is clear that interactive technologies allow incremental feedback that facilitates shared understanding and establishing common ground, as well as supporting incremental reference. Similarly, technologies that don't support interactivity (e.g. email, voicemail) show disrupted processes (e.g. reference), and outcome (lack of shared understanding). Technologies that transmit visual information about facial expressions and gaze may be critical for tasks in which affect and attitude have a central role. Technologies that don't support visual interpersonal information (e.g speech, email) lead to interactions that are more impersonal and likely to end in deadlock. Technologies that provide a shared visual environment are important when tasks require complex reference to, and joint manipulations of, physical objects.

These results also offer practical insights about why certain intuitively plausible novel technologies have largely failed. People have strong intuitions about the value of visual information about other people's gaze, gesture and facial expressions in communication (Isaacs & Tang, 1994; O'Conaill et al., 1993; Sellen, 1995). Nevertheless, counter to these intuitions, the cognitive cueing research indicates that such visual information isn't important either for turn-taking, or for availability. Yet many applications of video presuppose this type of non-verbal information (Whittaker, 1995). This may explain the failure of technologies that attempt to provide visual images of other conversational participants, (e.g. videoconferencing, videophone), when these technologies are used for cognitive tasks in work settings (Fish et al., 1992; 1993). Our research also suggests that video might be more profitably used to communicate information about shared objects in a shared environment rather than "talking heads" type applications (Anderson et al., 2000; Nardi et al., 1996; Whittaker, 1995; Whittaker & O'Conaill, 1997). The bandwidth hypothesis research also suggests reasons for the success of the phone. These results show the primacy of speech for a number of different tasks, along with demonstrating that combining other technologies with speech seldom improves performance, making it hard to improve on the phone (Chapanis

.
.
.

et al., 1972, 1977).

We can also make other practical recommendations about the use of various mediated communication technologies. Our results show that the effect of using mediated communication depends on the task. For social tasks there are clearly differences between mediated and face to face interaction, but for many cognitive tasks (especially those that don't require access to a shared physical environment) outcomes may not be different. Technology users may therefore decide to avoid (or at least be aware of the consequences) of using mediated communication if their current activity has a strong social component. Nevertheless, there are two problems in applying these types of guidelines directly. The first is the absence of a systematic task taxonomy (Straus & McGrath, 1994), enabling potential technology users to determine whether their current communication task is cognitive or social, prior to making their decision about whether to use technology for that task. A related point is that research on naturally occurring interactions shows that they are often heterogeneous: it is unusual for a single real-world communicative task to be predominantly social or cognitive, with most interactions containing elements of both (Short et al., 1976).

Mediated communication research also makes a theoretical contribution to theories of face to face communication. In one way, we can see technologies and their affordances as 'natural experiments', enabling us to dissect the contributions of different communication behaviours on communication processes, outcome and content. As we have noted, with the exception of work on nonverbal communication (Argyle, 1990; Beattie, 1978; 1981; Duncan, 1972; Kendon, 1967), most theories of face to face interaction have tended not to try to isolate the effects of different behaviours on communication process and content. In contrast the experiments we report here are extremely important in helping to clarify the role of visible behaviours (facial expressions, gaze, gesture). The data reviewed here show that visible behaviours such as gaze and gesture are important for communicating interpersonal information. However, contrary to people's intuitions and the claims of various studies (Argyle, 1990; Duncan, 1972; Kendon, 1967) they do not seem to be necessary for availability or turn-taking. The experiments reviewed here also document the role of interactivity, as being a vital communication process in mediating understanding. The studies of shared workspaces also provide evidence for the notion of grounding:

.
.
.

showing the effects of a shared visual environment on reference and explicitness of communication.

However, one weakness of many experiments on mediated communication has been the focus on technology instead of affordances, communication behaviours. Most experiments have treated technologies as independent variables. Although this provides important practical information for users of that technology, it does not enable us to theoretically disentangle the contributions of the technology's affordances on underlying behaviours. For example, many early experiments compared email with face to face communication, but when differences were discovered it was impossible to determine whether these resulted from the lack of visual information in email or the fact that it did not support interactivity. Future research needs to focus on underlying affordances and communication behaviours, rather than existing technologies per se.

What makes the picture more complex is that the causal relationships between affordances, communication behaviours and core communicative phenomena are highly intricate. A single affordance can support multiple communication behaviours, and these in turn can have effects on different core communicative phenomena. Thus the visual mode supports multiple behaviours (gaze, head nods, gesture, shared visual environment) relating to different aspects of communication (turn-taking, reference, attention, availability, interpersonal information). Similarly, one behaviour can affect multiple communication phenomena: backchannels support attention, reference, turn-taking, and agreement. Furthermore, communication behaviours are often redundant with respect to the core communication processes they affect. Thus gaze, gesture, head nods, and verbal backchannels all contribute to turn-taking. One consequence of all this is that it weakens predictions that can be made to test the effects of affordances on communication processes. For example removing visual information may have less serious effects on turn-taking than might be expected, because other communication behaviours that do not rely on the visual mode, namely verbal backchannels, tag questions, addressing, can substitute for the absence of visual information (O'Conaill et al., 1993, Whittaker & O'Conaill, 1997). This suggests that future research needs to be much clearer when testing hypotheses, about the relevant affordances, underlying behaviours and their anticipated effects on communication. To determine precise effects it may be necessary to configure

.
.
.

technologies to isolate different types of behaviours, e.g. to experiment with technologies that support the transmission of facial expressions, but not gaze, to determine the separate contributions of these behaviours.

Another theoretical problem is the lack of a clear underlying model of the relationship among core communicative phenomena such as processes, content, shared understanding and task outcome. Much of the cognitive cueing work presupposes a model in which technology affordances influence underlying communication behaviours, which then affect higher level processes such as understanding and conversational content (Clark, 1992; 1996; Clark & Brennan, 1991; McCarthy et al., 1991; 1993; O'Conaill et al., 1993; Oviatt & Cohen, 1991, Whittaker et al., 1991; 1993). In contrast, the social cueing work of Rutter (1987) and Short et al. (1976) proposes models in which technologies directly influence the content of communication which then affects communication processes and understanding in turn. According to these social cueing theories, technologies lacking the visual mode are more impersonal and formal, because of the absence of interpersonal content. This absence of affective information in turn affects higher level processes such as negotiation. Now while Rutter (1987) presents regression analyses supporting his particular causal model, our models generally need to be clearer about the relationship among technologies, affordances, processes, content and understanding.

There are a number of other limitations to our theoretical account. The first concerns the role of tasks. We have repeatedly observed that the effects of various technologies and their underlying affordances on communication depends on the task. The major differences occur between social and cognitive tasks. As noted above, however, we need to move beyond this simple distinction, and provide a richer task taxonomy. Although there have been repeated attempts to do this, in general they have not proved to be successful (Kraut et al., 1992, Straus & McGrath, 1994). Media richness theory (Daft & Lengel, 1984), for example, proposes that tasks can be categorized in terms of their equivocality (i.e. the existence of multiple or conflicting interpretations). However, the dimension of equivocality does not map directly onto existing taxonomies of group communications (McGrath, 1984, Straus & McGrath, 1994), nor does it correspond well with observations of naturally occurring workplace interactions (Short et al., 1976). Although other

.
.
.

task taxonomies have been proposed (McGrath, 1984, Morrison & Vogel, 1991), these can be subjected to both theoretical and empirical criticism (Short et al., 1976, Straus & McGrath, 1994). Further work on defining a task taxonomy is much needed.

Another fundamental limitation of our theories is the impoverished notion of affordances. An enriched taxonomy of affordances may address some of the problems observed above concerning the complex relations among affordances, behaviours and communication phenomena. It should mean that we are better able to map from technology affordances to predictions about communication. The current review has focused on two types of affordance, namely mode and interactivity. Although the simple distinction between mode and interactivity affordances explains much of the current data, the notion of affordances needs to be refined in a number of ways. First, the notion of visual information is too broad. We need to distinguish between aspects of visual information relating to affective information (e.g. facial expressions), turn-taking (e.g. gaze) and availability (e.g. physical presence). Similarly, it is apparent that rather than being a straightforward dichotomy, interactivity is a matter of degree. Numerous studies have shown that speech is more interactive than typing, although both are real-time technologies. Speech allows for precisely timed backchannels, interruptions and completions, and these processes that are much harder to achieve in typing (Oviatt & Cohen, 1991, Whittaker et al., 1991). Another underresearched affordance concerns the extent to which technologies support conversational *context*. One obvious property of many technologies, (e.g. voicemail, email, chat, instant messaging) is that they are permanent. Permanence means that prior conversational context is viewable and does not have to be remembered (Whittaker et al., 1991; McCarthy et al., 1991; 1993; Whittaker & Sidner, 1996; Whittaker, Swanson, Kucan & Sidner, 1997). Having context directly available has been shown to change conversational processes and outcome indicating the importance of this affordance (Whittaker et al., 1991; 1993; McCarthy et al., 1993). Another final affordance is *expressivity*. Comparisons of non-interactive uses of voice and text, for annotation (Kraut et al., 1992), and asynchronous communication (Whittaker, Hirschberg & Nakatani, 1998; Whittaker, Davis, Hirschberg & Muller, 2000), show that voice is used for different tasks than text. Voice is also rated by users as better for describing global and affective information than text. Together these data

.
.
.

suggest that we should enrich our set of affordances, to include richer definitions of visual information, interactivity, context and expressivity.

We also need to move away from purely comparative theories. Most mediated communication research explicitly uses models derived from face to face communication. Yet there are strong reasons to think that non-interactive communication raises a large number of problems that are not commonly encountered in face to face settings. Research into email and long term collaborations has documented the fact that non-interactive communications enable participants to participate in multiple simultaneous intermittent interactions (Kraut et al., 1993; Whittaker et al., 1994; 2000; Whittaker & Sidner, 1996; Whittaker, Jones & Terveen, 2002). Such interactions can take place sporadically over time intervals lasting from minutes to years. Furthermore, participants engage in multiple concurrent conversations of this type (Kraut et al., 1993; Whittaker et al., 1994; Whittaker & Sidner, 1996; Whittaker et al., 2002). This gives rise to a whole new set of theoretical and empirical questions about: (a) how participants monitor multiple concurrent conversations, (b) how they retain the context of those conversations over long time intervals, and (c) how they reinitiate an inactive conversation to regain the attention of their conversational partners. Long term intermittent conversations also raise important theoretical questions about common ground and shared understanding. Most face to face communication theories (e.g. Clark, 1992; 1996; Clark & Shaefer, 1989; Krauss et al., 1977) argue that shared understanding is achieved through interactive processes by which feedback can be given and misunderstandings quickly redressed. But in long term intermittent interaction such interactivity is absent, so how can discrepancies in understanding be detected and addressed and what stops misunderstandings from proliferating? These are important questions because they challenge fundamental assumptions about communication.

Another set of fundamental issues concerns the application of face to face theories to *mass communication*. Again, most mediated communication models are implicitly derived from a restricted set of circumstances: *dyadic* face to face communication. However, in many new forms of asynchronous communication involving technologies such as Usenet and Lotus Notes, interactions can take place between hundreds or even thousands of participants. Recent research has shown that the principles guiding

.
.
.

such mass interactions may be fundamentally different from face to face interaction, with longer, more involved discussions taking place between participants who have *less rather than more* common ground (Whittaker, 1996, Whittaker et al., 1998). This research argues that notions of *weak ties* rather than mutual understanding may better explain these data.

Together, these theoretical and empirical gaps call for more sophisticated theorizing about technology affordances and tasks. We also need more systematic research into mediated communication that does not presuppose face to face models. One strategy to achieve this might be to explore situations in which mediated communication is *better* than face to face communication (Hollan & Stornetta, 1992). For example, recent research has shown that in certain settings, instant messaging is preferred to face to face communication because it is less interruptive (Nardi et al., 2000). In conclusion, to move the field of mediated communication forward we need progress in a number of areas. We need to: refine our notions of task and the relations between tasks, to produce more careful analyses of important underlying explanatory affordances, to define new models that better characterize asynchronous communication and mass communication. Unless we do this, we will never have theories that truly address mediated communication, rather than trying to account for mediated communication results in terms of face to face communication.

## References

Abel, M. (1990). Experiences in an exploratory distributed organization. In J. Galegher, R. Kraut & C. Egido Eds. Intellectual Teamwork, pps 489-510. Hillsdale, N.J.: Lawrence Erlbaum Press.

Anderson, A.H., Bard, E., Sotillo, C., Doherty-Sneddon, G., & Newlands, A. (1997). The effects of face-to-face communication on the intelligibility of speech. Perception and Psychophysics, 59, 580-592.

Anderson, A.H., Smallwood, L., MacDonald, R., Mullin, J., Fleming, A. & O'Malley, C. (2000) Video data and video links in mediated communication: what do users value?, *International Journal of Human-Computer Studies*, 52(1), 165-187**.**

Argyle, M. (1990). Bodily communication London, Routledge.

Argyle, M., & Cook, M. (1976). Gaze and mutual gaze. Cambridge, Cambridge University Press.

Argyle, M. and Graham, J. (1977). The Central Europe Experiment - looking at persons and looking

. . .

at things. Journal of Environmental Psychology and Nonverbal Behaviour, 1, 6-16.

Argyle, M., Lalljee, M., & Cook, M. (1968). The effects of visibility on interaction in the dyad. Human Relations, 28, 289-304.

Argyle, M., Lefebvre, L., & Cook, M. (1974). The meaning of five patterns of gaze. European Journal of Social Psychology, 4, 125-136.

Beattie, G. (1978). Sequential Patterns Of Speech And Gaze In Dialogue. Semiotica, 23, 29-52.

Beattie, G. (1981). A Further Investigation Of The Cognitive Interference Hypothesis Of Gaze Patterns In Conversation. British Journal of Social Psychology, 20, 243-248.

Bly, S. (1988). A use of drawing surfaces in collaborative settings. In Proceedings of Conference on Computer Supported Cooperative Work, 250-256, New York: ACM Press.

Bly, S., Harrison, S., & Irwin, S. (1993). Media spaces: Bringing people together in a video, audio and computing environment, Communications of the ACM, 36, 28-45.

Bruce, V. (1995). The Role Of The Face In Face-To-Face Communication: Implications For Videotelephony. In S. Emmot (Ed.), Information superhighways: multimedia users and futures. London, Academic Press.

Chapanis, A., Ochsman, R., Parrish, R., & Weeks, G. (1972). 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.

Chapanis, A., Ochsman, R., Parrish R., and Weeks, G. (1977). Studies in interactive communication. II: The effects of four communication modes on the linguistic performance of teams during cooperative problem solving, Human Factors, 19, 487-509.

Chesebro, J. (1985). Computer mediated interpersonal communication. In B. Ruben, Ed., Information and Behavior, 1, 202-224. New Brunswick, NJ: Transactions

Clark, H. (1992). Arenas of language use. Chicago: University of Chicago Press.

Clark, H. (1996). Using language. Cambridge University Press.

Clark H. & Brennan, S. (1991). Grounding in communication. In L.B. Resnick, J. Levine & S.

.
.
.

Teasley, Eds. Perspectives on socially shared cognition. Washington DC., APA Press.

Clark H. & Marshall, C. (1981). Definite reference and mutual knowledge. In A. Joshi, B. Webber & I. Sag Eds. Elements of discourse understanding. Cambridge, Cambridge University Press.

Clark, H., & Schaefer, E. (1989). Contributing to discourse. Cognitive Science, 13, 259-292.

Clark, H. & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. Cognition, 22, 1-39.

Cohen, K. (1982). Speaker interaction: video teleconferences versus face-to-face meetings. In Proceedings of teleconferencing and electronic communications, 189-199. Madison: University of Wisconsin Press.

Cooper, R. (1974). The control of eye fixation by the meaning of spoken language. Cognitive Psychology, 6, 84-107.

Culnan, M. J., & Markus, M.L. (1987). Information technologies. In F.M. Jablin, L.L. Putman, K.H. Roberts, & L.W. Porter (Eds.), Handbook of organizational communication: An interdisciplinary perspective. Thousand Oaks: Sage.

Daft, R. and Lengel, R. (1984). Information richness: a new approach to managerial behavior and organizational design. In B. Straw and L. Cummings, Eds. Research in Organizational Behaviour, pps 191-233, Greenwich, CT., JAI Press.

Daly B. (1993) The influence of face to face versus computer mediated communication channels on collective induction. Accounting, Management, and Information Technologies, 3, 1-22.

Daly-Jones, O., Monk, A., & Watts, L. (1998). Some advantages of video conferencing over high quality audio conferencing: fluency and awareness of attentional focus, International Journal of Human-Computer Studies. 49, 21-58.

Doherty-Sneddon, G., Anderson, A., O'Malley, C. , Langton, S. , Garrod, S., and Bruce, V. (1997) Face-to-face and video mediated communication: a comparison of dialogue structure and task performance. Journal of Experimental: Psychology (Applied), 3, 105-125.

Dubrovsky, V., Kiesler, S., and Sethna, B. (1991). The equalization phenomenon: status effects in computer mediated and face-to-face decision making groups. Human Computer Interaction, 6, 119-146.

.
.
.

Duncan, S. (1972). Some signals and rules for taking speaker turns in conversation. Journal of Personal and Social Psychology, 23, 283-292.

Ekman, P., (1982) Emotions in the Human Face. Cambridge: Cambridge University Press.

Ekman, P. & Friesen, W. (1975). Unmasking the face. New Jersey: Prentice Hall.

Exline, R., & Winters, L. (1965). Effects Of Cognitive Difficulty And Cognitive Style On Eye Contact In Interviews. In Proceedings of the Eastern Psycholoical Association, Atlantic City, NJ, 35-41.

Gibson, J. J. (1966). The senses considered as perceptual systems. Allan and Unwin, London.

Fish, R., Kraut, R., & Chalfonte, B. (1990). The videowindow system in informal communications. In Proceedings of Conference on Computer Supported Co-operative Work, 1-12, New York: ACM Press.

Fish, R., Kraut, R., Root, R. & Rice, R. (1992). Evaluating video as a technology for informal communication, In Proceedings of CHI'92 Human Factors in Computing Systems, 37-48, New York: ACM Press.

Fish, R., Kraut, R., Root, R. & Rice, R. (1993). Video as a technology for informal communication. Communications of the ACM. 36, 48-61.

Fulk, J., Schmitz, J., and Steinfield, C. (1990). A social influence model of technology use. In J. Fulk and C. Steinfield Eds., Organizations and communication technology, Newbury Park, CA: Sage.

Gaver, W., Sellen, A, Heath, C., & Luff, P. (1993). One is not enough: multiple views in a media space. In Proceedings of CHI'93 Human Factors in Computing Systems, 335-341, New York: ACM Press.

Heath, C. and Luff, P., Disembodied Conduct: Communication through video in a multi-media environment, Proc. CHI'91, 99-103, New York, ACM Press.

Hiltz, S.R. (1975). Communications and group decision making: experimental evidence on the potential impact of computer conferencing (Research Report 2, Newark, NJ: New Jersey Institute of Technology Computerized Conferencing and Communications Center.

Hiltz, S.R., Johnson, K., & Agle, G. (1978). Replicating Bales problem-solving experiments on a computerized conferencing system, Report 8. Newark, NJ: New Jersey Institute of Technology Computerized Conferencing and Communications Center.

.
.
.

Hiltz, S. R., Johnson, K., & Turoff, M. (1986). Experiments in group decision making: Communication process and outcome in face-to-face versus computerized conferences. Human Communication Research, 13, 225-252.

Hiltz, S. R., & Turoff, M. (1978). The network nation: Human communication via computer. Reading, MA: Adison-Wesley.

Hiltz, S. R., Turoff, M., & Johnson, K. (1989). Experiments in Group decision making, 3: Disinhibition, deindividuation, and group process in pen name and real name computer conferences. Decision Support Systems, 5, 217-232.

Hollan, J. and Stornetta, S. (1992). Beyond being there. In Proceedings CHI'92 Conference on Human Factors in Computing Systems, Perspectives on the Design of Collaborative Systems, 119-125, New York: ACM Press.

Hopper, A. (1994). Communications at the desktop. Computer Networks and ISDN Systems, 26, 1253-1265.

Isaacs, E., & Tang, J. (1994). What video can and can't do for collaboration: a case study. MultiMedia Systems, 3, 63-73, Anaheim., CA.

Isaacs, E., Whittaker, S., Frohlich, F., and O'Conaill, B. (1997). Informal communication re-examined: New functions for video in supporting opportunistic encounters". In K. Finn, A. Sellen, and S. Wilbur, Eds., Video-mediated communication, pps 459-486. Mahwah, NJ: Lawrence Erlbaum Associates, 1997.

Jaffe, J., and Feldstein, S. 1970. Rhythms of Dialogue. New York.

Jackson, M., Anderson, A., McEwan, R., and Mullin, J. (2000). Impact of video frame rate on communicative behaviour in two and four party groups. In Proceedings of Conference on Computer Supported Cooperative Work, ACM, New York.

Kahneman, D. (1973). Attention and effort. New Jersey: Prentice Hall.

Kelly, J., Futoran, G., and McGrath, J. (1990). Capacity and capability: seven studies of entrainment of task performance in task performance rates, Small Group Research, 21, 283-314.

.
.
.

Kelly, J. and McGrath, J. (1985). Effects of time limits and task types on task performance and interaction of four person groups. Journal of Personality and Social Psychology, 49, 395-407.

Kendon, A. (1967). Some functions of gaze direction in social interaction. Acta Psychologica, 26, 1-47.

Kendon, A. & Ferber, A. (1973). A Description Of Some Human Greetings. In R. Michael & J. Crook (Eds)., Comparative ecology and behaviour of primates, pps 591-668. London, Academic Press.

Kiesler, S., Siegel, J., and McGuire, T. (1984). Social psychological effects of computer mediated communication, American Psychologist, 39, 1123-1134.

Kiesler, S., Zubrow, D., Moses, A. M., & Geller, V. (1985). Affect in computer-mediated communications: An experiment in synchronous terminal-to-terminal discussion. Human-Computer Interaction, 1, 77-104.

Kleck, R. & Nuessle, W. (1968). Congruence Between The Indicative And Communicative Functions Of Eye-Contact In Interpersonal Relations. British Journal of Social and Clinical Psychology, 7, 241-246.

Kleinke, C. (1986). Gaze and eye contact: a research review. Psychological Bulletin, 100, 78-100.

Krauss, R. & Bricker, P. (1967). Effects of transmission delay and access delay on the efficiency of verbal communication. Journal of the Acoustical Society of America, 41, 286-292.

Krauss, R. M., Bricker, P. D., McMahon, L. E., & Garlock, C. M. (1977). The role of audible and visible back channel responses in interpersonal communication. Journal of Personality and Social Psychology, 35, 523-529.

Krauss, R. M., & Weinheimer, S. (1966). Concurrent feedback, confirmation and the encoding of referents in verbal communication. Journal of Personality and Social Psychology, 4, 343-346.

Kraut, R., Fish, R., Root, B. & Chalfonte, B. (1993). Informal communication in organizations. In R. Baecker Ed., Groupware and Computer Supported Cooperative Work, pps 287-314. San Mateo, California: Morgan Kaufman.

Kraut, R. Galegher, J., Fish, R., and Chalfonte, B. (1992). Task requirements and media choice. Human Computer Interaction, 7, 375-407.

.
.
.

Kraut, R., Lewis, S., and Swezey, L. (1982). Listener responsiveness and the coordination of conversation. Journal of Personality and Social Psychology, 43, 718-731.

Kraut, R., Miller, M., and Siegel, J. (1996). Communication in performance of physical tasks: effects on outcomes and performance. In Proceedings of the Conference on Computer Supported Cooperative Work, 57-66, New York: ACM Press.

Lea, M. & Spears, R. (1991). Computer-mediated communication, de-individuation and group decision-making. International Journal of Man-Machine Studies, 24, 283-301.

McCarthy, J.C., Miles, V.C. & Monk, A.F. (1991). An experimental study of common ground in text-based communication. In S. Robertson, G. Olsen, & J. Olsen, (Eds.), Proceedings of CHI91 Conference on Computer Human Interaction, 209-215, New York: ACM Press.

McCarthy, J.C., Miles, V.C., Monk, A.F., Harrison, M.D., Dix, A.J. & Wright, P.C. (1993). Text-based on- line conferencing: a conceptual and empirical analysis using a minimal prototype, Human Computer Interaction, 8, 147-184.

McLeod, P. (1992). An assessment of the experimental literature on electronic support of group work: Results of a meta-analysis. Human Computer Interaction, 7, 257-280.

McGrath, J. (1984). Groups: Interaction and Performance, New Jersey: Prentice-Hall.

McGuire, T. W., Kiesler S., & Siegel J. (1987). Group and computer-mediated discussion effects in risk decision-making. Journal of Personality and Social Psychology, 52, 917-930.

Mantei, M, Baecker, R, Sellen, A, Buxton, W, Milligan, T, & Wellman, B. (1991). Experiences in the use of a media space. In Proceedings of CHI'91 Human Factors in Computing Systems, 203-209, New York: ACM Press.

Markus, L. (1987). Toward a Critical Mass Theory of Interactive Media: Universal Access, Interdependence, and Diffusion, Communication Research, 14, 491-511.

Markus, L. (1994). Electronic mail as the medium of managerial choice, Organizational Science, 5, 502-527.

Meyers, D. (1985). Home computer communication networks. New Orleans: Loyola University

.
.
.

Department of Communications.

Morley, I., and Stephenson, G. (1969). Interpersonal and interparty exchange. A laboratory simulation of an industrial negotiation at the plant level. British Journal of Psychology, 60, 543-545.

Morley, I.E. & Stephenson, G.M. (1970). Formality in experimental negotiations: a validation study. British Journal of Psychology, 61, 383- 384.

Morrison, J., and Vogel, D., (1991). Group support from a task based perspective, Support Functionality in the Office Environment, 1, 3-22.

Nardi, B., Kuchinsky, A., Whittaker, S., Leichner, R., and Schwarz, H. (1996). Video as data: technical and social aspects of a collaborative multimedia application. In Computer Supported Cooperative Work, 4, 73-100.

Nardi, B., Schwarz, H., Kuchinsky, A., Leichner, R., Whittaker, S. & Sclabassi, R. (1993). Turning away from talking heads: an analysis of "video-as-data". In Proceedings of CHI'93 Human Factors in Computing Systems, 327-334, New York: ACM Press.

Nardi, B., Whittaker, S., Bradner, E. (2000). Interaction and Outeraction: Instant Messaging in Action. In *Proceedings of Conference on Computer Supported Cooperative Work,* 79-88*,* ACM, New York.

Nardi, B., Whittaker, S., and Schwarz, H. (2000). It's Not What You Know, It's Who You Know: Work in the Information Economy. *First Monday*, May 2000. www. firstmonday.org.

Nardi, B., Whittaker, S., Schwarz, H. (in press). A networker's Work is Never Done: Joint Work in Intensional Networks. To appear in Journal of Computer-supported Cooperative Work.

Norman, D. (1988). The psychology of everyday things Basic Books, New York.

Norwhine, A., and Murphy, O. (1938). Characteristic time intervals in telephonic conversations. Bell System Technical Journal, 17, 281-291.

O'Conaill, B., Whittaker, S., & Wilbur, S. (1993). Conversations over videoconferences: an evaluation of the spoken aspects of video mediated interaction. Human Computer Interaction, 8, 389-428.

Olaniran, B. A. (1994). Group performance in computer-mediated and face-to-face communication

.
.
.

media. Management Communication Quarterly, 7: 256-281.

Oviatt S, and Cohen P. (1991) Discourse Structure and Performance Efficiency in Interactive and Non-Interactive Spoken Modalities. Computer Speech and Language, 5, 297-326.

Olson, J.S., Olson, G.M., & Meader, D.K. (1995) What mix of video and audio is useful for remote real-time work? Proceedings of CHI '95, 362-368, New York, ACM Press.

Reid, A. (1977). Comparing the telephone with face-to-face interaction. In I. Pool Ed., The Social Impact of the Telephone, pps 386-414. Cambridge, MA: MIT Press.

Reid, F., Ball., L., Morley, A., Evans, J. (1996). Styles of group discussion in computer mediated decision making. British Journal of Social Psychology, 36, 241-262.

Rice, R. E. (1984). Mediated group communication. In R. Rice (Ed.). The new media (pp. 129-154). Newbury Park: Sage.

Rice, R. (1993). Media appropriateness: using social presence theory to compare traditional and new organizational media, Human Communication Research, 19, 451-484.

Rice, R. and Love, G. (1987). Electronic emotion: socioemotional content in a computer mediated network. Communication Research, 14, 85-108.

Rice, R. & Shook, D. (1990). Voice messaging, coordination and communication. In J. Galegher, R. Kraut, & C. Egido (Eds.) *Intellectual teamwork: Social and technological bases of cooperative work*. (pp. 327-350.) New Jersey: Erlbaum.

Root, R. (1988). Design of a multi-media vehicle for social browsing. In Proceedings of Conference on Computer Supported Co-operative Work, 25-38, New York: ACM Press.

Rutter, M, (1987). Communicating by telephone. Oxford: Pergamon Press.

Rutter, D., & Robinson, R. (1981). An experimental analysis of teaching by telephone. In G. Stephenson & J. Davies (Eds.), Progress in applied social psychology, London: Wiley Press.

Sellen, A. (1992). Speech patterns in video-mediated communication. In Proceedings of CHI'92 Human Factors in Computing Systems, 49-59, New York: ACM Press.

Sellen, A. (1995). Remote Conversations: The Effects of Mediating Talk with Technology. Human

.
.
.

Computer Interaction, 10, 401-441.

Schegloff, E. (1982). Discourse as an interactional achievement: Some uses of "uh huh" and other things that come between sentences. In D. Tannen, Ed., Analyzing discourse: text and talk, pps 71-93. Washington: Georgetown University Press.

Short, J. (1971). Conflicts of interest and conflicts of opinion in an experimental bargaining game conducted over the media. Communication Studies Group Report. London: University College.

Short, J. (1972). Effects of medium of communication on experimental negotiation. Human Relations, 27, 225-234.

Short, J, Williams E, and Christie, B. (1976). The social psychology of telecommunications. London: Wiley Press.

Siegel, J., Dubrovsky, V., Kiesler, S., and McGuire, T. (1986). Group processes in computer mediated communication, Organizational Behavior and Human Decision Processes, 37, 157-187.

Sproull, L. (1983). The nature of managerial attention. In P. Larkey and L. Sproull (Eds.), Advances in Information Processing in Organizations, Greeenwich, Ct: JAI Press.

Sproull, L. and Kiesler, S. (1986). Reducing social context cues: electronic mail in organizational communication. Management Science, 32, 1492-1512.

Sproull, L. and Kiesler, S. (1991). Connections: new ways of working in the networked organization. Cambridge, MIT Press.

Stefik, M., Foster, G., Bobrow, D., Kahn, K., Lanning, S., and Suchman, L., (1987). Beyond the Chalkboard: Computer Support for Collaboration and Problem Solving in Meetings, *Communications of the ACM*, 30(1), 32-47.

Stephenson, G., Ayling, K., and Rutter, D. (1976). The role of visual communication in social exchange. British Journal of Social and Clinical Psychology, 15, 113-120.

Straus, S. G. & McGrath, J. E. (1994). Does the medium matter? The interaction of task type and technology on group performance and member reactions. Journal of Applied Psychology, 79, 87-97.

Steinfield, C. (1986). Computer mediated communication in an organizational setting: Explaining

. . .

task-related and socio-emotional issues. In M. McLaughlin, Ed., Communication Yearbook 9, 777-804, Newbury Park, CA: Sage.

Tang, J., (1991) Findings from Observational Studies of Collaborative Work, International Journal of Man-Machine Studies, 34, 143-160.

Tang, J. & Isaacs, E. (1993). Why do users like video: studies of multimedia-supported collaboration. Computer Supported Cooperative Work, 1, 163-196.

Tang, J., Isaacs, E, & Rua, M. (1994). Supporting distributed groups with a Montage of lightweight interactions . In Proceedings of Conference on Computer Supported Co-operative Work, 23-34, New York: ACM Press.

Tang, J. & Rua, M. (1994). Montage: Providing teleproximity for distributed groups. In Proceedings of CHI'94 Human Factors in Computing Systems, 37-43, New York: ACM Press.

Tatar, D., Foster, G., & Bobrow, D. (1991). Design for Conversation: Lessons from Cognoter, *International Journal of Man-Machine Studies*, 34(2), 185-210.

Walker, M., and Whittaker, S. (1990). Mixed Initiative in Dialogue: An investigation into discourse segmentation . In Proceedings. 28th Annual Meeting of the Association for Computational Linguistics, 70-76, Cambridge, MA, MIT Press.

Walther, J. (1992). A longitudinal experiment on relational tone in computer mediated and face to face interaction. In J. Nunamaker and R. Sprague, Eds., Proceedings of the Hawaii Interactional Conference on Social System Sciences, 4, 220-231, Los Alamitos, CA: IEEE Press.

Walther, J. (1994). Anticipated ongoing interaction versus channel effects on relational communication in computer mediated interaction, Human Communication Research, 20, 473-501.

Walther, J. (2002). Time effects in computer mediated groups. To appear in P. Hinds & S. Kiesler, (Eds.), Distributed work, pps 235-258. Cambridge, MA: MIT Press.

Watt, R. (1995). An Examination Of The Visual Aspects Of Human Facial Gesture. In S. Emmot (Ed.), Information superhighways: multimedia users and futures. London, Academic Press.

Weisband, S. (1992). Group discussion and first advocacy effects in computer-mediated and face-to-

.
.
.

face decision making groups. Organizational Behavior and Human Decision Processes, 53, 352-380.

Whittaker, S. (1995) Rethinking video as a technology for interpersonal communications: theory and design implications. International Journal of Man-Machine Studies, 42, 501-529.

Whittaker, S. (1996). Talking to strangers: an evaluation of the factors affecting electronic collaboration. In Proceedings of Conference on Computer Supported Cooperative Work, 1996, 409-418, New York: ACM Press.

Whittaker, S., Brennan, S., & Clark, H. (1991). Coordinating activity: an analysis of computer supported co-operative work. In Proceedings of CHI'91 Human Factors in Computing Systems, 361-367, New York: ACM Press.

Whittaker, S., Davis, R., Hirschberg, J., and Muller, U. (2000). Jotmail: a voicemail interface that enables you to see what was said. In *Proceedings of CHI2000 Conference on Human Computer Interaction*, 89-96, ACM: New York.

Whittaker, S., Frohlich., & Daly-Jones, O. (1994). Informal workplace communication: what is it like and how might we support it? In Proceedings of CHI'94 Human Factors in Computing Systems, pp. 130-137, New York: ACM Press.

Whittaker, S., Geelhoed, E., & Robinson, E. (1993). Shared workspaces: how do they work and when are they useful? International Journal of Man-Machine Studies, 39, 813-842.

Whittaker, S., Hirschberg, J., and Nakatani, C. H. (1998). All talk and all action: strategies for managing voicemail messages, In *Proceedings of CHI98 Conference on Computer Human Interaction,* 1998.

Whittaker, S., Jones, Q., and Terveen, L. (2002). Managing Long Term Conversations: Conversation and Contact Management. In Thirty-fifth Annual Hawaii International Conference on System Sciences, (Jan 7-10th Big Island Hawaii), 427-437.

Whittaker, S. & O'Conaill, B. (1993). Evaluating videoconferencing. In Companion Proceedings of CHI'93 Human Factors in Computing Systems, New York: ACM Press.

Whittaker, S, & O'Conaill, B. (1997). The role of vision in face-to-face and mediated communication.

.
.
.

In K.E. Finn, Sellen, A.J. et al. (Eds.). Video-mediated communication: Computers, cognition, and work, 23-49. Mahwah, N.J.: Lawrence Erlbaum.

Whittaker, S. and Sidner, C. (1996). Email overload: exploring personal information management of email. In Proceedings of CHI'96 Conference on Computer Human Interaction, 276-283, New York: ACM Press.

Whittaker, S. & Stenton, P. (1988). Cues and control in expert client dialogues. In Proceedings of the Conference for the Association for Computational Linguistics, 123-130, Cambridge, MA.: MIT Press.

Whittaker, S., Swanson, G., Kucan, J., & Sidner, C., (1997). Telenotes: managing lightweight interactions in the desktop. Transactions on Computer Human Interaction, 4, 137-168.

Whittaker, S. Terveen, L., Hill, W., and Cherny, L. (1998). The dynamics of mass interaction, In Proceedings of the Conference on Computer Supported Cooperative Work, 257-264, New York: ACM Press.

Williams, E. (1977). Experimental comparisons of face-to-face and mediated communication. Psychological Bulletin, 16, 963-976.

Yngve, V. (1970). Getting A Word In Edgewise. In Proceedings of the Sixth Meeting of the Chicago Linguistics Society, 567-577, Chicago, Chicago Linguistic Society.