DISCOURSE PROCESSES, *31*(1), 91–110

Language Use in Computer-Mediated Communication: The Role of Coordination Devices

Jeffrey T. Hancock
Department of Psychology
Dalhousie University
Halifax, Canada

Philip J. Dunham
Department of Psychology
Dalhousie University
Halifax, Canada

In text-based computer-mediated communication (CMC), people's language use and task performance can be affected by whether explicit coordination devices (e.g., a turn marker) are available. Analyses of 37 dialogues that emerged in text-based, CMC environments support Clark's (1996, 1997) proposal that a communication setting that disrupts the regulation of turn-taking will both undermine higher level language processes (i.e., the construal of meaning) and increase the frequency of metacommunicative signals required to coordinate the speaker's action with the listener's attention. The results indicate that the availability of a simple, explicit turn marker in a task-oriented, text-based exchange facilitates the construal of meaning and reduces the number of verbal coordination devices required to ground communication. Measures examining alternative interpretations indicate that these effects are not easily explained by differences in volubility or speed–accuracy trade-off across the 2 conditions.

In his most recent analysis of language use, Clark (1996, 1997) outlined a comprehensive theoretical framework that conceptualizes discourse broadly as two or more participants engaged in the collaborative process of building a mutually agreed on common ground, comprised of the beliefs, attitudes, expectations, and other information presupposed by participants to be mutual knowledge (Clark, 1996; Clark & Schaefer, 1989). In this article, our goal is to explore several im-

---

Correspondence and requests for reprints should be sent to Philip J. Dunham, Department of Psychology, Dalhousie University, Halifax, Nova Scotia, Canada B3H 4J1. E-mail: dunham@is.dal.ca

plications of Clark's analysis of language use in the context of the more impoverished communicative environment of computer-mediated communication (CMC).

In Clark's (1996) model of human communication, common ground was assumed to emerge from two different sources. The first consists essentially of all the beliefs, knowledge, and presuppositions believed to be held in common by all participants in the exchange. The second source of common ground is information emerging during the interaction itself (Krauss & Fussell, 1991). The collaborative emergence of common ground during conversation assumes that all participants eventually reach the mutual belief that what a speaker meant "has been understood by everyone well enough for current purposes" (Clark & Schaefer, 1989, p. 290). The process of reaching this mutual belief has been referred to as *grounding* (Clark, 1996; Clark & Brennan, 1991).

In this broad conceptual framework, Clark's (1996) model defined several levels of communicative acts in which the interactive grounding process is assumed to operate. Adopting the metaphor of a *joint action ladder*, the rungs on the ladder define different dimensions of communicative behavior that are in progress during any single slice of time. More specifically, four hierarchically organized levels of language use are assumed important when characterizing the acts of both a speaker and listener (i.e., joint actions). All four levels of communicative behavior must be coordinated successfully for new information to accumulate in common ground. A general *principle of upward causality* is also assumed to guide this collaborative process by requiring interlocutors to coordinate lower levels (e.g., turn-taking) prior to grounding higher levels of language use (e.g., joint construal of meaning). For example, at the lowest level of the joint action ladder, conversants typically must coordinate the listener's attention with the speaker's communicative actions (Level 1). If the speaker executes a communicative act while the listener's attention is otherwise engaged, the grounding process may be undermined. Indeed, the emergence of a turn structure in conversation is portrayed as one consequence of the speaker and listener synchronizing their communicative activities at this lowest rung of the joint action ladder (Brennan, 1990; Clark, 1994).

At the second level, once the communicative act has been successfully executed and attended to, the listener must attempt to identify signals presented by the speaker (e.g., an English word, a gesture, etc.). Higher rungs on the action ladder are concerned with what the listener understands the speaker to mean (Level 3) and ultimately with the completion of projects or actions proposed in the conversation (Level 4). For example, during a dinner conversation the speaker and listener must succeed in jointly construing the meaning of a speaker's utterance "Please pass the salt" at Level 3 before a listener can consider the speaker's proposal to pass the salt (Level 4). Essentially, the model assumes that the speaker and listener must ground their actions at each level of this hierarchically organized joint action ladder or risk failing to contribute to common ground.

When problems arise during this collaborative process, people employ several strategies to restore communication. Clark's (1996) model made an important distinction, for example, between primary presentations, utterances that are signals about the official business or the purpose of the discourse (i.e., *Track 1 signals*), and secondary presentations, signals that are about the communicative acts themselves (i.e., *Track 2 signals*). Track 2 signals serve to repair the grounding process at each level of the joint action ladder. For example, during face-to-face conversations, at the lowest level of the joint action ladder people use a variety of techniques, such as initiating side sequences (e.g., "What?" or "Hold on a minute") and nonverbal behavior (e.g., gaze, gestures, etc.), to synchronize and regulate their activities and turn-taking (Clark, 1997; Duncan, 1972; Jefferson, 1972). Similarly, at higher levels on the joint action ladder, conversants exploit Track 2 signals to repair miscontruals of meaning (e.g., quizzical expressions, verbal utterances "Whadaya mean?" etc.). Note that in both examples the metacommunicative, Track 2 signals are about the communication itself and not the official business of the conversation.

Obviously not all human communication occurs in the rich, dynamic face-to-face settings presupposed by Clark's (1996, 1997) collaborative model. For example, we talk on the telephone, we write letters, and we employ various other technologies to communicate. To the degree that various settings impoverish or enrich the different levels of the communicative context, Clark and his colleagues have acknowledged that the collaborative grounding process should be affected by various constraints imposed by the communication medium (Clark & Brennan, 1991). Indeed, the constraints imposed across various media provide the researcher with some interesting opportunities to explore the various mechanisms described (e.g., McCarthy, Miles, & Monk, 1991; Whittaker, 1995). For example, the sound-only context of traditional telephony eliminates, among other things, the interlocutors' ability to monitor facial expressions and gestures, reducing potentially useful information available in face-to-face conversations (Boyle, Anderson, & Newlands, 1994; Rutter, 1987). Strategies that have emerged to compensate for constraints in telephony include the summons–answer sequence identified by Schegloff (1968) in which the answerer speaks first in identification or greeting and the caller then identifies him or herself and provides the first topic of conversation. The summons–answer sequence has presumably emerged in telephone conversations to facilitate the coordination of attention and action (i.e., Level 1 of Clark's joint action ladder) in this medium.

Similar questions are being raised in the rapidly growing literature concerned with CMC. The increased use of networked computers in contemporary society for both synchronous text-based exchanges (e.g., Internet chat groups) and for asynchronous exchanges of e-mail has generated a considerable amount of interest in exploring the impact of this novel medium on language use (Chapanis & Overbey, 1974; Clark & Brennan, 1991; Ferrara, Brunner, & Whittemore, 1991; Garcia & Jacobs, 1999; Herring, 1996; McCarthy & Monk; 1994; McCarthy, Wright, & Monk, 1992; McKinlay, Proctor, Masting, Woodburn, & Arnott, 1994; Whittaker, Brennan, & Clark, 1991; Williams, 1977). Much of the existing data attempts to document differences in CMC and face-to-face language use. When differences are observed, inferences are generally made about the nature of constraints imposed by the technology. For example, Ferrara et al. reported that conversants in a text-based synchronous communicative environment tend to omit copulas, subject pronouns, and articles. They argue that this strategy is a response to the increased costs of formulating and producing utterances (i.e., typing is more difficult than speaking). Similarly, McCarthy et al. observed that during text-based exchanges participants tended to develop topics in parallel and to address the intended listener more explicitly in an effort to maintain coherence during text-based exchanges.

This research was designed to extend work concerned with the constraints that CMC imposes on language use and was prompted by some interesting data suggesting that participants may find it particularly difficult to coordinate turn-taking in text-based conversations that occur in real time (Beattie, 1986; Beattie, Cutler, & Pearson, 1982; McKinlay et al., 1994). Text-based exchanges in the CMC environment preclude the use of many of the signals considered important for coordinating turn-taking during face-to-face interactions, such as eye gaze (Duncan & Fiske, 1977; Goodwin, 1981; Kendon, 1967), gestures and nonverbal behavior (Bavelas, Chovil, Coates, & Roe, 1995; Duncan & Fiske, 1977), and nonlinguistic vocal behaviors (e.g., back channel utterances, prosody, pitch, etc.; Beattie, 1986; Duncan, 1972; Yngve, 1970). Beattie et al. (1982), for example, found that when most of these signals were eliminated in a text-only format, people were unable to judge reliably whether utterances (transcribed from a television interview) had occurred in the middle or at the end of the speaker's actual turns. These participants were, however, quite accurate when watching a video or hearing an audio recording of the same interview. In addition, important cues like pause length, although available in the synchronous CMC environment, are distorted relative to face-to-face (McKinlay et al., 1994).

If, as these data suggest, the coordination of turns is difficult in real-time text-based environments, when viewed from the perspective of Clark's (1996) joint action ladder, the deleterious effects on the grounding process should be serious. Recall that problems associated with the coordination of turns reflects a disruption of the grounding process at the lowest level of the joint action ladder. According to the principle of upward causality, disruptions at this level should undermine the grounding process at higher levels of communication (e.g., joint construal of meaning), and participants encountering difficulties in coordinating communicative activities should more frequently display Track 2 signals to repair and facilitate the grounding process.

To assess the extent to which difficulty in managing turn allocation in the CMC environment has an impact on the construction of common ground during dyadic conversation, we directly compared conversations occurring in two different CMC environments. In one environment, two individuals exchanged typewritten messages in real time using a What You See Is What I See (WYSIWIS) interface. Within this communicative environment the speaker's messages are continually streamed onto the addressee's computer screen, character by character. The screen, which is divided into separate windows for each participant's messages, simultaneously displays both participants' text. Because the messages can be continuously streamed and participants are physically isolated, there are few inherent turn-taking signals to inform conversants when to type and when to read (i.e., participants can type messages simultaneously). To facilitate subsequent discussion, a CMC interface with these generic properties will be described as an Unmarked CMC environment.

In the second environment, participants communicated through essentially the same interface but with an explicit turn marker added (e.g., a keystroke designating the end of a message), or in other words a marked CMC environment. As Clark and Brennan (1991) noted, the presence of an explicit turn marker, such as "o" for over, in the latter interface should facilitate the coordination of the speaker's acts with the addressee's attention. In both environments, dyads exchanged messages in an attempt to match a series of abstract shapes (i.e., tangrams) that provided a clearly defined and objective measure of mutual understanding (see Schober & Clark, 1989).

Although previous CMC research has focused on either outcome measures (e.g., task performance; McKinlay et al., 1994) or process measures (e.g., Garcia & Jacobs, 1999; for a review of this issue, see McCarthy & Monk, 1994), random assignment of participants to the two conditions of this controlled experimental procedure permitted us to examine both outcome-based as well as process-related implications of Clark's (1996) model in this impoverished communicative environment. First, errors on the tangram-matching task measured the degree to which the dyads were successful in grounding higher levels of communication (e.g., Level 3, understanding and meaning). Second, the joint action ladder analysis implies that participants encountering difficulty in synchronizing their activities in the Unmarked CMC environment should employ Track 2 signals regulating attention and action more frequently than participants in the Marked CMC environment.

METHOD

Participants

Participants were 80 English-speaking (24 men and 56 women) members of the Dalhousie University community who received course credit or token remuneration for participating in the experiment. Participants were randomly paired to form 40 dyads. The members of each dyad were not acquainted with each other prior to the experiment. Participants knew when they volunteered that a partner would be involved. More women volunteered to participate in the study; consequently, random pairings of participants produced 3 male–male, 18 male–female, and 19 female–female combinations. Three dyads (6 participants) were excluded from analysis because 1 of the participants in the dyad was not a native English speaker and experienced difficulty communicating in this context.

Procedure

Skill assessments. Prior to participation, each individual answered a series of questions designed to assess levels of relevant prior computer experience. Three dichotomous categories were used to identify individual differences in competence: (a) regular (i.e., more than once per week) versus infrequent (i.e., less than once per week) computer use; (b) regular versus infrequent access to newsgroups or e-mail (i.e., less than once per week); and (c) real-time chat experience versus no experience. In addition, following participation in the procedure, individual differences in typing skills were assessed online with a 1-min test using Ainsworth Typing Trainer 3 package. Finally, all participants completed the Neo-Five Factor Personality Inventory (Costa & McCrae, 1991). The latter measure was part of a different project conducted in parallel with this research.

Communication task. Participants used one of two desktop computer stations located in separate rooms while the experimenter monitored and recorded the interaction from a third station. The three computers were networked such that signal transmission was virtually instantaneous between stations. Each terminal was equipped with the Windows95™ operating system and Mirabilis ICQ software (v. Beta1.113). The video displays were split into two regions so that participants could view their own text and that of their partners (i.e., using the Horizontal option of the ICQ package). Partners could send and receive messages simultaneously, with messages displayed on a WYSIWIS or character-by-character basis. Note that this type of text-based chat, in which messages are composed in public, differs from other systems in which messages are composed in private and sent to a public message area (e.g., Internet Relay Chat). The WYSIWIS system contains no inherent turn signals, unlike Internet Relay Chat systems, which delimit the beginning and end of a turn when a message "unit" is transmitted.

On arrival at the laboratory, the experimenter escorted each participant separately to their respective computer stations, avoiding any contact between the members of a dyad prior to the task. Once participants were situated at their terminals, the experimenter briefly demonstrated the use of the computer interface. Both participants were then told that they would collaborate in text-based exchanges in order to complete a figure-matching task. The task was adapted from Schober and Clark (1989). One participant, randomly designated the Director, described a series of tangrams to another participant, the Matcher, and the latter attempted to identify the tangrams being described. During the task, the Director was given three sheets of paper with four numbered tangrams printed on each sheet (see Figure 1). The Director was instructed to describe each tangram, one at a time, to the Matcher. Each tangram was discussed until the Matcher claimed to have identified it.

The Matcher was also provided with three sheets of paper. There was an array of 16 tangrams and 4 “tangram not present” choices on each sheet (see Figure 2). The Matcher’s task was to search the array in an effort to identify each tangram being described by the Director. The Matcher was encouraged to ask questions during this process and instructed to notify the Director when he or she believed the tangram had been correctly matched. Once they had identified a particular tangram, the Matcher was instructed to write the identifying number beside the corresponding tangram.

In addition, participants were cautioned that tangrams may be reversed in orientation (left/right) and that to make the task challenging some tangrams would be missing, although in fact none were. If the Matcher concluded that the tangram was missing, the Matcher was to label one of the “tangram not present” boxes with the appropriate identifying number. Both Director and Matcher were instructed not to reveal their name, age, or sex during the interaction.

Dyads participating in the tangram task were randomly assigned to either the Marked or Unmarked communicative environment. In the Marked condition, dyads were instructed to use the letter “o” (short for “Over”) to signal the completion of their messages (Clark & Brennan, 1991). This provided an explicit mechanism for coordinating turn allocation. An analysis of the transcripts revealed that participants in the Marked condition failed to mark their turn completions on only 1.2% of the turns. On these rare occasions, the experimenter immediately reminded them to correctly mark their turn completions. In the Unmarked condition, no explicit signal was designated to regulate turn allocation. Once the procedure was completed, each member of the dyad was introduced to

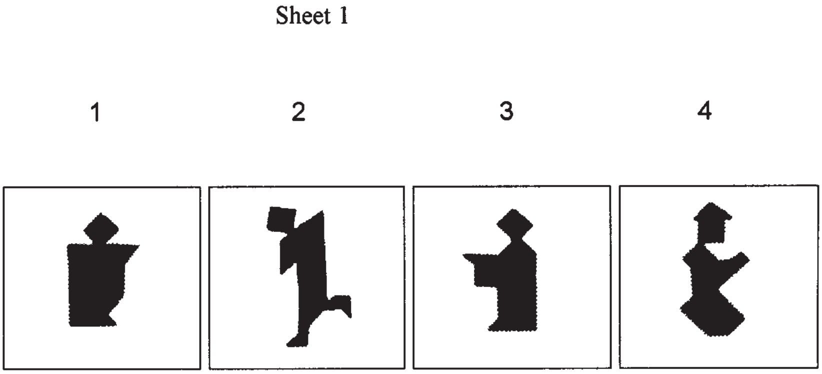

FIGURE 1 Tangram sheet 1 for the Director. The Director was instructed to describe the tangrams one at a time in sequence according to their number.

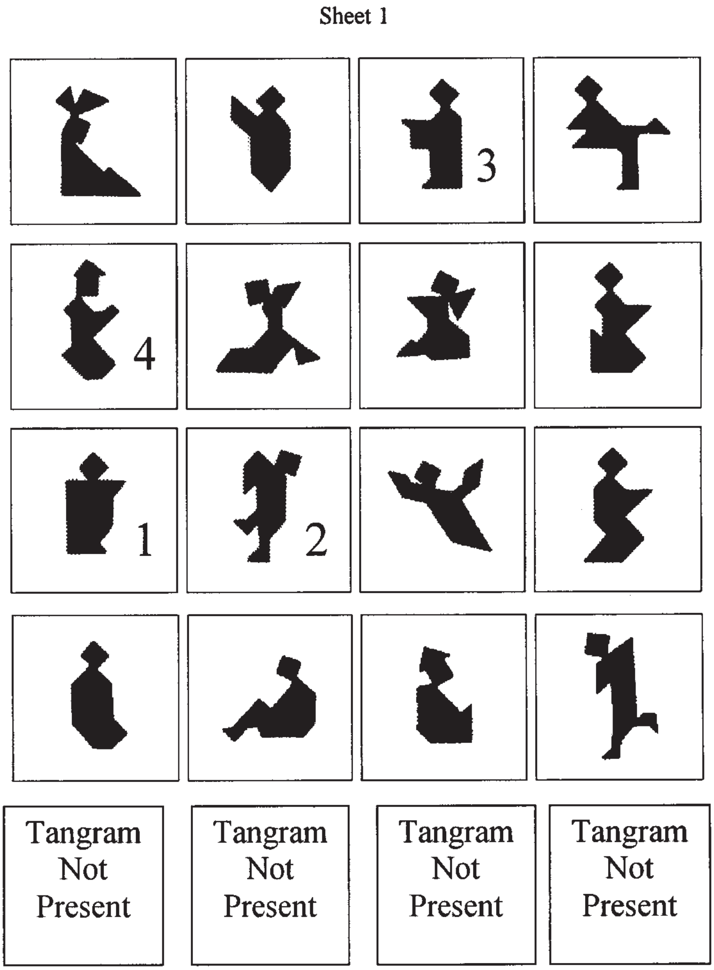

FIGURE 2 Tangram sheet 1 for the Matcher. Numbers within the boxes are examples of the identifying number inserted by the Matcher according to the order in which the Director described the tangrams. In this example, Tangram 2 has been misidentified.

his or her partner and debriefed. Although some participants recognized each other during the debriefing phase, no participants knew who their partner was during the task.

Dependent measures. First, the frequency of tangram identification errors made by dyads in each treatment condition were counted as an index of the extent to which common ground failed to develop at the higher levels of language use (i.e., joint construal of meaning; see Schober & Clark, 1989). An error was defined as a mismatch between the tangram described by the Director and the tangram identified by the Matcher. The maximum number of errors possible was 12. We also measured the total time each dyad spent completing the tangram-matching task (i.e., identify all 12 tangrams), and the transcripts of all interactions were examined to determine the total number of words exchanged by the participants. The number of words in the transcript, omitting deletions and replacements, as well as a separate file that included all keystrokes was recorded for each dyad. Additionally, the average number of words per turn was calculated for each condition. In the Unmarked Condition, turns were based on adjacency-pair units (see next), in which each half of an adjacency pair was counted as one turn. In the Marked Condition, turns were delimited by the "o" signal.

Finally, the frequency of Track 2 coordination devices employed by dyads in each treatment condition was also counted. Track 2 signals associated with the execution and attention to communicative acts (the lowest level on the Joint Action Ladder) were coded from the transcripts of the conversations. To identify these Track 2 signals, a coding scheme first identified messages as belonging to one of three phases for each tangram-matching dialogue: the Presequence, Body, or Exit phase (Clark, 1996). The Presequence phase consisted of utterances concerned with moving from one tangram to the next and preparing for tangram description. The Body phase consisted of all utterances concerned with the description and identification of the tangram. The Exit phase consisted of utterances that confirmed a match. Because, by definition, the Presequence and Exit phases of the tangram dialogues were concerned with the negotiation of the tangram task (i.e., agreeing to progress from one tangram to the next), these phases were not considered in the Track 2 analysis. Only messages within the Body phase of each tangram-matching dialogue, which do not have the additional coordination burden of task negotiation, were coded for the presence of Track 2 coordination devices. When coding for Track 2 devices, the unit of analysis was one half of an adjacency pair. Adjacency pairs, as defined by Schegloff and Sacks (1973), consist of two ordered utterances or parts produced by different speakers. Given a first pair part, the second pair part of an adjacency pair is conditionally relevant or expectable as the next utterance. Either component of the adjacency pair could be coded as a Track 1 or Track 2 utterance. Only those utterances containing signals concerned with the coordination of the interaction, the lowest level of the joint-action ladder (e.g., "Give me a sec," "OK, continue"), were included in the Track 2 analysis.

Coding reliability. Jeffrey T. Hancock coded Track 2 measures for all 37 dyads. A second observer, trained on the rating system, rated 5 random dyads from each treatment condition (total of 10 dyads). Cross-tabulation of the frequency data from each independent observer at the most detailed level of the coding scheme (i.e., tracks and levels of talk) yielded a chance-corrected kappa statistic of .90 (Bakeman & Gottman, 1986).

RESULTS

Recall that a higher percentage of correctly matched tangrams is assumed to index more successful grounding at higher levels of the joint-action ladder (Schober & Clark, 1989). These data revealed, as predicted, that dyads provided with an explicit turn completion signal in the Marked environment were more accurate on the tangram task (86.3% correct; $SE = 2.78$) than dyads in the Unmarked environment (76.3% correct; $SE = 4.37$), $t(35) = 2.02$, $p < .05$, $\eta^2 = .11$. Refer to Table 1 for examples of tangram descriptions that resulted in errors in the Unmarked and Marked conditions.

Also, as predicted, dyads in the Unmarked communicative environment employed Track 2 coordination devices more frequently ($M = 5.2$, $SE = 1.68$) than dyads in the Marked communicative environment ($M = 0.88$, $SE = 0.44$), $t(35) = 2.39$, $p < .05$, $\eta^2 = .13$; $t(21.6) = 2.57$, $p < .05$, when corrected for unequal variances. This observation is further confirmed by the greater proportion of dyads in the Unmarked environment who employed Track 2 signals (13 dyads employed at

TABLE 1
Examples of Descriptions That Lead to Errors in the Unmarked and Marked Conditions

Unmarked condition	
Dyad UM129, Sheet 2, Tangram 3	
Director	This one looks like a house that is angled on its left corner that is your left corner. Just infront of that there is a straight line.
Matcher	i don't think that i have this figure on my sheet.
Director	This lline continues downward to about 45 degrees simialr to an arrow.
Matcher	i have nothing that looks like a house that I tilted!
Matcher	the only possibility is that if the figure appears to have an arrow head pointing downward.
Director	The figure is wearing a triangle on its head.]
Matcher	is this still #3 or haave we moved on?
Director	yes we are still on figure 3
Matcher	i think that i don't have number 3 because nothing on my sheet resembles what you are discribing. It is possible that the tangram doesn't exist
Marked condition	
Dyad M135, Sheet 1, Tangram 2	
Director	if this were again a person the person is leaning towarrds the left with his left shoulder pointing down. his right shoukder is pointing up and his right leg or foot is lifted in the air o
Matcher	is his head a diamond shape?o
Director	yes o
Matcher	i am still unclear about his right shoulder it looks to be underneath the left or vice versa, if I am looking at the right picture o
Director	yes his right shoulder is only a point where his top of his shoulder and his side of the body meet o
Matcher	I think I have it, tangram 2 now finished o

Note. These examples do not represent how the utterances appeared on screen.

least one Track 2 signal) in comparison to the Marked environment (5 dyads employed at least one Track 2 signal), $\chi^2(1, N = 37) = 4.66, p < .05, \eta^2 = .13$. Several qualitative examples of Track 2 Level 1 signals from the Unmarked condition transcripts are displayed in Table 2 and from the Marked condition in Table 3.

To determine whether the tendency to employ Track 2 signals or utterances was related to the errors made by a dyad, the frequency of these utterances was correlated with the total number of errors across dyads in the Marked and Unmarked conditions. Within the Unmarked environment this correlation was significant, $r = -.48, p < .05$, suggesting that those dyads who actively managed communicative action and attention were also most successful at the higher levels of communication, as indexed by tangram errors. This same correlation was not significant in the Marked environment, $r = .09$, *ns*. However, when interpret-

TABLE 2
Examples of Track 1 and Track 2 (Level 1) Signals in the Unmarked Condition

Speaker	*Track 1*	*Track 2*
Dyad UM119, Sheet 1, Tangram 1		
Director	and is zig- zag going down.	
Director		Any questions?
Matcher		hold on a sec, let me think
Matcher	i got it	
Dyad UM109, Sheet 2, Tangram 1		
Director	. . . parallel w/ the ground /his/her seat	
Director		did that help?
Matcher		hold on
Director		k
Matcher		okay, continue
Dyad UM111, Sheet 1, Tangram 2		
Director		Questions?
Matcher		hold on I am looking
Matcher	what do the other parts look like?	
Director		While you are looking, I'll describe the other parts!
Dyad UM111, Sheet 2, Tangram 1		
Director	[. . .] and you can't see the hands or arms	
Director		Anything?
Matcher		checking it out
Dyad UM113, Sheet 2, Tangram 4		
Director	diamond head	
Director		still there?
Matcher		YEAH
Dyad UM131, Sheet 3, Tangram 3		
Director	(types 9 lines of text)	
Matcher		wow. Just a minute
Director		k :)
Matcher	are his "arms" triangles?	

TABLE 3
Examples of Track 1 and Track 2 (Level 1) Signals in the Marked Condition

Speaker	*Track 1*	*Track 2*
Dyad M111, Sheet 1, Tangram 1		
Director	a triangle. with only one point visible.o	
Matcher	ok o	
Director		you have identified him? o
Matcher		can I ask one more question? o
Director		ya sure, o
Dyad M109, Sheet 2, Tangram 1		
Director	the last image is my personal favorite.. Are you a figure skating fan? This one is Kurt Browning. 0	
Matcher	I do love Kurt!	Let me look o
Director		k, u look, I'll continue..
	he's got two armts out, on one leg... o	

Note. Track 2 Level 1 signals were rare in the Marked Condition (15 signals total).

ing the latter association, it should be noted that the absolute frequency of Track 2 signals was low, and variance was consequently restricted.

To assess the extent to which participants adapted over the session to the constraints imposed by these two text-based communicative environments, we also looked for any change in the frequency of errors and frequency of Track 2 devices across the first and second half of the session. A 2 (marked vs. unmarked) × 2 (first six tangrams vs. last six tangrams) analysis of variance (ANOVA) revealed no significant difference in errors across the first and second half of the session, and no interaction between this repeated measures factor and the Marked vs. Unmarked environments (both F values < 1). A similar analysis on the frequency of Track 2 devices revealed no significant difference in the use of this metacommunicative device across the first and second half of the session ($F < 1$). Essentially, both analyses reveal that the participants' scores were not changing systematically across this single experimental session.

Ancillary Measures

Several additional measures suggest that the differences in tangram task performance are not easily explained by differences in volubility across the two treatment conditions. An examination of the transcripts revealed that the total number of words exchanged did not differ between the Unmarked ($M = 961$, $SE = 115.3$) and Marked ($M = 1012$, SE = 142.8) interfaces, $t(35) = .29$, *ns*, and neither did the total number of keystrokes generated per dyad (Unmarked condition: $M = 5697.34$, $SE = 739.36$; Marked condition: $M = 5853.35$, $SE = 798.33$), $t(34) = -.14$, *ns*. Furthermore, correlational analyses revealed at best, a minimal, nonsignificant negative relation between the number of words exchanged and number of errors on the task ($r = -.22$, *ns*).

An examination of the adjacency pair units also revealed that the mean frequency of turns generated per dyad did not differ significantly across conditions (Unmarked Condition: $M = 71.20$, $SE = 8.52$; Marked Condition: $M = 61.24$, $SE = 4.24$), $t(35) = .991$, *ns*. Similarly, although the Marked Condition had slightly more words per turn ($M = 17.50$, $SE = 3.27$) than the Unmarked Condition ($M = 13.95$, $SE = 1.24$), this difference was also not significant, $t(35) = 1.08$, *ns*.

The total amount of time taken (in minutes) to complete the tangram task per dyad was recorded and varied widely within treatment conditions (Unmarked: $M = 36.2$ min, $SE = 4.13$; Marked: $M = 41.2$, $SE = 4.87$) but did not differ across conditions, $t(35) = .79$, *ns*. To reduce any potential influence of outliers in these data, a nonparametric analysis was also conducted on the times to completion across the two conditions and yielded the same result (Mann–Whitney $U(1) = 141$, *ns*). Once again, the correlation between the time to complete the task and the number of tangram errors was minimal, negative, and nonsignificant ($r = -.25$, *ns*). Considered together, these latter data indicate that differences in construal of meaning (i.e., tangram errors) observed across these two environments do not involve some form of speed–accuracy trade-off.

In terms of computer skills, all but 1 participant in this study used computers on a regular basis (i.e., at least once per week), and most used computers for e-mail or newsgroup communication (Directors: $M = 86.5\%$; Matchers: $M = 89.2\%$). Experience with Chat programs was less common (Directors: $M = 43.2\%$; Matchers: $M = 37.8\%$) but did not differ across treatment conditions, $\chi^2(1, N = 37) = .44$, *ns*, and was not significantly related to tangram task performance (Directors: $r = .17$, *ns*; Matchers: $r = -.03$, *ns*).

Because the task was collaborative in nature, it is possible that previous chat experience would be valuable only when both participants had experience with the medium. To assess this possibility, a total dyadic computer experience score was calculated by summing the individual scores to produce a “dyadic” experience score (i.e., neither participant has prior chat experience = 0; one participant has chat experience = 1; both participants have chat experience = 2). Again, there was no significant relation between the dyadic chat experience score and tangram errors ($r = .12$, *ns*). Accordingly, individual differences in prior computer experience were not considered further.

Although typing skill has not typically been assessed in computer-mediated communication research, it is conceivable that typing speed has an impact on turn coordination or task completion times, or both. Word per minute (wpm) and percentage correct accuracy scores were therefore measured for each participant based on a 1-min typing test. Participants’ net wpm was calculated by multiplying their gross wpm by their percentage accuracy. Individual typing speeds did not vary across roles (Directors: 31.9 wpm; Matchers: 31.8 wpm). The dyad’s net wpm was also calculated by averaging the two participants’ net wpm. Dyadic typing ability (Unmarked condition: $M = 29.67$ wpm; Marked condition: $M = 34.17$ wpm) again did not differ across conditions, $t(35) = -1.45$, *ns*, and these measures did not correlate with tangram task performance (Unmarked condition: $r = -.04$, *ns*; Marked condition, $r = -.08$, *ns*), Track 2 signals ($r = .04$, *ns*) or time to task completion ($r = .04$, *ns*). Despite its potential to affect turn-taking protocol, typing ability was not significantly related to tangram performance or Track 2 expression and was not considered further.

DISCUSSION

In this research, we explored two basic implications of Clark's (1996, 1997) model of language use in the novel setting of text-based, computer-mediated exchanges. First, we tested his prediction that disruptions in the coordination of lower level communicative acts (e.g., joint attention and turn structure) should undermine the grounding process at higher levels of communication (e.g., construal of meaning). Second, we explored the assumption that interlocutors will tend to employ metacommunicative Track 2 signals to the degree that the grounding process has been disrupted at any particular level. Results from this study provided support for both of these predictions in the real-time, text-based exchanges of a CMC environment.

A more detailed analysis of the discourse produced by the dyads suggests that the superior performance in the Marked environment was associated with improved coordination of communication at the level of joint-attentional processes and turn allocation. Several sources of evidence converge on this conclusion. First, dyads in the Unmarked condition more frequently employed Track 2 signals than dyads in the Marked condition. Second, a larger proportion of the dyads in the Unmarked condition displayed Track 2 signals than in the Marked condition. Finally, and perhaps most importantly, within the Unmarked condition, dyads who implemented Track 2 coordination devices made fewer errors on the tangram task than those who failed to employ Track 2 signals.

These differences and this inverse relation between tangram errors and Track 2 signals ($r = -.48$) provide strong evidence for the principle of upward causality assumed in Clark's model. Higher level processes required to match these abstract figures were specifically dependent on the efficient use of lower level coordination devices required to establish joint attention. It is also important to recognize that the quantity of information and task completion times did not differ across conditions, nor did the frequency or average length of turns. As such, the effects are not easily explained by differences in volubility or a speed–accuracy trade-off. Considered together, the findings indicate that the differences observed at higher levels of communication (i.e., tangram performance) appear to be influenced by modifications in language use at lower levels of the joint action ladder (i.e., explicit turn markers and ad hoc Track 2 coordination devices).

Given the difference in task outcome observed in this study, some additional speculation about the nature of the processes that may underlie these differences in performance seems warranted. First, we note that Clark's (1996) assumption that the coordination of lower level language processes was an important prerequisite for the success of higher level comprehension clearly has considerable intuitive appeal in the context of a face-to-face conversation. During face-to-face exchanges, communicative actions are typically evanescent and if not attended to with precise timing are lost to the continuous, collaborative process of constructing common ground (Clark, 1996). Although the data from this experiment are compatible with this argument, we also believe that Clark's reasoning loses some of its intuitive appeal in the context of text-based, computer-mediated communication. Text-based, CMC exchanges are obviously less evanescent than actual speech, and participants can potentially refer back to the conversational record after messages have been exchanged. Indeed, given continuous access to the text of a CMC interaction, one might argue that it should be relatively easy to repair any negative consequences of lower level failures to coordinate. Why then, in the data presented here, does the disruption of lower level coordination processes in the Unmarked condition also disrupt higher level processes in this more permanent text-based medium?

Although this experiment was not designed to address the importance of evanescence directly, a closer look at the conversations in the Unmarked condition suggests that the presence of prior information on the computer screen is not sufficient to prevent disruption of the grounding process. Consider, for example, the following breakdown in joint-construal of meaning observed in the Unmarked condition (please note that the following example does not represent how the utterances appeared on the screens):

Director: okay tangram #1
looks kinda like a short person . . . with a diamond
(square on its tip) for a head
the rest of the body . . . well . . . it's easier if I describe
what's missing than what's there. . .
it's just a big rectangle (tall, not long) with a section
missing from its right side
Matcher: ok, got it
Director: on the right side, there's a rectangle with a equilateral
triangle MISSING (so the shoulder, if it were one,
comes to a point.
got it . . . meaning got the tangram?
Matcher: yes, i believe so.

In the previous example, the Director assumed that the Matcher was attending to her communicative action and provided three brief descriptive statements in succession while the Matcher inserted a confirmation (“ok, got it”) after the Director’s third statement. Because of her incorrect assumption regarding the Matcher’s attention status, the Director did not attend to the Matcher’s confirming statement and proceeded with yet a fourth descriptive statement, reflecting a disruption in the first level of the joint-action ladder (Clark, 1996). When the Director eventually attended to the Matcher’s confirming statement, the grounding process at Level 3 of Clark’s model (i.e., meaning and understanding) had also been disrupted, as evidenced by the Director’s attempt to clarify the meaning of the Matcher’s confirming claim (“got it, meaning got the tangram?”).

Clearly, in the Unmarked condition, in the absence of any Track 2 signals, the Director found it difficult to monitor the Matcher’s attentional status at any given point of time. Indeed, when the Track 2 coordination devices are categorized by speaker and according to their function in the interaction (see Table 4), 82.7% of the Track 2 utterances signaled by the Director were concerned with inquiring about the Matcher’s current attentional status. These particular coordination devices are similar in both structure and function to “tag” questions observed in communicative environments, such as videoconferencing, requiring more formal techniques for transferring conversational initiative (e.g., O’Conaill, Whittaker, & Wilbur, 1993; Sellen, 1995). Additionally, 100% of the signals made by Matchers provided evidence about the Matcher’s current attentional status, either by directing the Director to halt or continue communicative activity (e.g., “hold on” “continue”) or by informing the Director about their current activity (e.g., “still looking” “checking”).

Consistent with the collaborative nature of Clark’s model of language use, the example described previously illustrates that even when prior records of the conversation are available on the screen, the absence of attentional coordination can weaken or disrupt the comprehension of a simple anaphoric reference such as “ok, got it.” In the Marked Condition, in contrast, once the Director has officially ended her turn with the signal “o,” the Director knows that she must wait until the Matcher responds. In fact, additional contributions, in which the Director added extra information after ending a turn but before the Matcher responded, were rare in the Marked condition (only 2.4% of turns in this condition constituted second installments). Consequently, the Matcher was able to shift his attention

TABLE 4
Taxonomy of Track 2 Signals According to Speaker and Function

Role	*Inquire*[a]		*Direct*[b]		*Inform*[c]		*Total*	
Director	62	(82.7%)	13	(17.3%)	0	(0%)	75	(70.1%)
Matcher	0	(0%)	20	(62.5%)	12	(37.5%)	32	(29.9%)
Total	62	(57.9%)	33	(32.4)	12	(11.2%)	107	

[a] See it? Still there? [b] Hold on; Continue. [c] Looking; checking it out.

to the tangram array without the risk of missing additional utterances. That is, the turn signal in the Marked Condition functioned to synchronize the interactants' task-based activities with their communicative activities.

When this example of the confusion that can arise during exchanges in the Unmarked CMC environment is considered along with the data described earlier, it seems reasonable to suggest that, even with a permanent text-based record, the absence of traditional turn markers or a turn protocol (Goodwin, 1981; Sacks, Schegloff, & Jefferson, 1974; see Garcia & Jacobs, 1999, for a discussion of Sacks et al.'s turn-taking analysis in this context) substantially increases the cognitive burden and effort that Clark (1996; Clark & Brennan, 1991) described as necessary to maintain the grounding process. Although both the Marked and the Unmarked CMC environments in this study severely restricted the kinds of cues typically used for lower level coordination in face-to-face interaction (Beattie, 1986; McKinlay et al., 1994; Rutter & Stephenson, 1977), only the Marked condition provided an explicit device designed to reduce additional competition for cognitive resources imposed by a more deliberate management of lower level coordination (cf. Norman & Shallice, 1986). Although speculative and a question for future research, in our view this increased competition for cognitive resources is perhaps a factor in the inferior performance observed in the Unmarked CMC environment.

A second point of interest in these data concerns the negative correlation between Track 2 metacommunicative signals and tangram errors in the Unmarked communicative environment. As noted, some participants dealt with the constraints of the Unmarked condition by employing ad hoc Track 2 signals to coordinate the speaker's action with the listener's attention, and this improved their tangram performance (see Table 2). For example, one Director consistently ended statements with the query "Any questions?" An analysis of this dyad's transcript revealed that in each instance the Director waited for the Matcher's response after signaling "Any questions?" Thus, this Director's signal regulated the interaction by indicating that the Matcher had the Director's attention and should at that point respond. The Director's consistent use of this tag question was functionally equivalent to the explicit turn-signal "o" in the Marked condition and coordinated the action and attention process at Level 1 of Clark's (1996) model, and presumably also helped to reduce any additional cognitive burden imposed by the Unmarked condition.

These observations are also consistent with previous findings demonstrating the ability of humans to develop adaptive strategies that help to overcome constraints imposed by various impoverished communicative environments (e.g., Boyle et al., 1994; Clark & Brennan, 1991; Garcia & Jacobs, 1999; Herring, 1999; McCarthy et al., 1992; Schegloff, 1968). For example, McCarthy et al. observed that participants interacting in small groups in a text-based communicative environment often used explicit verbal cohesive markers (e.g., "as a means of") and compressed messages in an attempt to hold the speaker role and maintain coherence. Similarly, Herring more recently described several forms of adaptation, including backchannels, turn signals, and cross-turn references, observed in Internet chat rooms. Indeed, some validity for the manipulation in this study is provided by one user, for example, who reported to Herring that

> he and a regular IRC conversational partner had devised the practice of typing a '%' symbol at the end of a message to indicate that they were not yet ready to give up the floor, and thus that the other should wait before taking the next turn.

Clearly, by adjusting their interaction style (e.g., increased verbal cohesive markers, increased Track 2 signals), interlocutors are able to establish coherence and ground their communication in text-based settings lacking typical coordination mechanisms, such as prosody, gaze, and deictic expressions.

Finally, in conclusion, we believe these data also illustrate that CMC environments have considerable potential for exploring various models of social interaction and language use. This impoverished communicative environment can be manipulated in a number of ways that directly test the role of various structures (e.g., coordination devices) presumed to be important in human social communication. We also believe the results of such manipulations will contribute substantially to our understanding of underlying mechanisms (e.g., coordination of action and attention) and to our understanding of this novel medium per se.

ACKNOWLEDGMENTS

This article is based on Jeffrey T. Hancock's master's thesis, supervised by Philip J. Dunham. This research was supported by Social Sciences and Humanities Research Council of Canada Grants 410–97–0763 awarded to Philip J. Dunham. We express our appreciation to Frances Dunham for technical support and to Diana Chard and Aimee Hancock for assistance with coding.

REFERENCES

Bakeman, R., & Gottman, J. (1986). *Observing interaction: An introduction to sequential analysis.* Cambridge, England: Cambridge University Press.

Bavelas, J. B., Chovil, N., Coates, L., & Roe, L. (1995). Gestures specialized for dialogue. *Personality-and-Social-Psychology-Bulletin, 21*, 394–405.

Beattie, G. (1986). *Talk: An analysis of speech and non-verbal behaviour in conversation.* Stony Stratford, England: Open University Press.

Beattie, G., Cutler, A., & Pearson, M. (1982). Why is Mrs. Thatcher interrupted so often? *Nature, 300*, 744–747.

Boyle, E. A., Anderson, A. H., & Newlands, A. (1994). The effects of visibility on dialogue and performance in a cooperative problem solving task. *Language & Speech, 37*, 2–20.

Brennan, S. E. (1990). *Seeking and providing evidence for mutual understanding.* Unpublished doctoral dissertation, Stanford University, California.

Chapanis, A., & Overbey, C. M. (1974). Studies in interactive communication: III. Effects of similar and dissimilar communication channels and two interchange options on team problem solving. *Perceptual and Motor Skills, 38*, 343–374.

Clark, H. H. (1994). Discourse in production. In M. A. Gernsbacher (Ed.), *Handbook of Psycholinguistics* (pp. 985–1022). San Diego, CA: Academic.

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

Clark, H. H. (1997). Dogmas of understanding. *Discourse Processes, 23*, 567–598.

Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. In L. B. Resnick, J. M. Levine, & S. D. Teasley (Eds.), *Perspectives on socially shared cognition* (pp. 127–149). Washington, DC: American Psychological Association.

Clark, H. H., & Schaefer, E. F. (1989). Contributing to discourse. *Cognitive Science, 13*, 259–294.

Costa, P. T., & McCrae, R. R. (1991). *NEO-FFI Form S.* Odessa, FL: Psychological Assessment Resources.

Duncan, S. (1972). Some signals and rules for taking speaking turns in conversations. *Journal of Personality and Social Psychology, 23*, 283–292.

Duncan, S., & Fiske, D. W. (1977). *Face-to-face interaction.* Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Ferrara, K., Brunner, H., & Whittemore, J. (1991). Interactive written discourse as an emergent register. *Written communication, 8*, 8–34.

Garcia, A. C., & Jacobs, J. B. (1999). The eyes of the beholder: Understanding the turn-taking system in quasi-synchronous computer-mediated communication. *Research on Language and Social Interaction, 32*, 337–367.

Goodwin, C. (1981). *Conversational organization: Interaction between speakers and hearers.* London: Academic.

Herring, S. C. (Ed.). (1996). *Computer-mediated communication: Linguistics, social and cross-cultural perspectives.* Philadelphia: Benjamins.

Herring, S. C. (1999). Interactional coherence in CMC. *Journal of Computer-Mediated Communication* [On-line], *4*. Retrieved August 3, 1999, from the World Wide Web: http://www.ascusc.org/jcmc/vol4/issue4/herring.html

Jefferson, G. (1972). Side sequences. In D. Sudnow (Ed.), *Studies in social interaction* (pp. 294–338). New York: Free Press.

Kendon, A. (1967). Some functions of gaze direction in social interaction. *Acta Psychologia, 26*, 22–63.

Krauss, R. M., & Fussell, S. R. (1991). Constructing shared communicative environments. In L. B. Resnick, J. M. Levine, & S. D. Teasley (Eds.), *Perspectives on socially shared cognition* (pp. 172–200). Washington, DC: American Psychological Association.

McCarthy, J. C., Miles, V. C., & Monk, A. F. (1991). An experimental study of common ground in text-based communication. *CHI'91 Conference Proceedings*, 217–215.

McCarthy, J. C., & Monk, A. F. (1994). Measuring the quality of computer-mediated communication. *Behaviour & Information Technology, 13*, 311–319.

McCarthy, J. C., Wright, P. C., & Monk, A. F. (1992). Coherence in text-based electronic conferencing: Coupling text and content. *Journal of Language and Social Psychology, 11*, 267–277.

McKinlay, A., Proctor, R., Masting, O., Woodburn, R., & Arnott, J. (1994). Studies of turn-taking in computer-mediated communication. *Interacting with computers, 6*, 151–171.

Norman, D. A., & Shallice, T. (1986). Attention to action: Willed and automatic control of behavior. In R. J. Davidson, G. E. Schwartz, & D. Shapiro (Eds.), *Consciousness and self-regulation: Advances in research and theory* (Vol. 4, pp. 1–18). London: Plenum.

O'Conaill, B., Whittaker, S., & Wilbur, S. (1993). Conversations over video conferences: An evaluation of the spoken aspects of video-mediated communication. *Human-Computer Interaction, 8*, 389–428.

Rutter, D. R. (1987). *Communicating by telephone.* Oxford, England: Pergamon.

Rutter, D. R., & Stephenson, G. (1977). The role of visual information in synchronizing conversation. *European Journal of Social Psychology, 2*, 29–37.

Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking in conversation. *Language, 50*, 696–735.

Schegloff, E. A. (1968). Sequencing in conversational openings. *American Anthropologist, 70*, 1075–1095.

Schegloff, E. A., & Sacks, H. (1973). Opening up closings. *Semiotica, 8*, 289–327.

Schober, M. F., & Clark, H. H. (1989). Understanding by addressees and overhearers. *Cognitive Psychology, 21*, 211–232.

Sellen, A. J. (1995). Remote conversations: The effects of mediating talk with technology. *Human-Computer Interaction, 10*, 401–444.

Whittaker, S. (1995). Rethinking video as a technology for interpersonal communications: Theory and design implications. *International Journal of Human-Computer Studies, 42*, 501–529.

Whittaker, S., Brennan, S. E., & Clark, H. H. (1991). Coordinating activity: An analysis of interaction in computer-supported co-operative work. In *Proceedings of ACM CHI'91 Conference on Human Factors in Computing Systems* (pp. 361–367). New York: Association for Computing Machinery Press.

Williams, E. (1977). Experimental comparisons of face-to-face and mediated communication: A review. *Psychological Bulletin, 84*, 963–976.

Yngve, V. H. (1970). On getting a word in edgewise. In *Papers from the sixth Regional Meeting of the Chicago Linguistic Society* (pp. 567–578). Chicago: Chicago Linguistic Society.