To cite, please refer to: Fox Tree, J. E., Mayer, S. A., & Betts, T. E. (2011). Grounding in instant messaging. *Journal of Educational Computing Research,* 45(4) 455-475.

Grounding in Instant Messaging

Jean E. Fox Tree, Sarah A. Mayer, Teresa E. Betts

University of California Santa Cruz

Author Note

This research was supported by faculty research funds granted by the University of California Santa Cruz. We thank our many research assistants who aided in data collection and coding, with a special thanks to Amanda Gutwirth, Allison Osborn, Jared Gorley, Stephanie Crockett, Erin Carter, Meg Grounds, and Flo Capinding. We also give a special thanks to Jack Tomlinson who assisted in getting the project off the ground. Correspondence can be addressed to Jean E. Fox Tree, Psychology Department, Social Sciences II room 277, University of California, Santa Cruz, CA, 95064, USA, foxtree@ucsc.edu, phone: 831-459-5181; fax: 831-459-3519.

Abstract

In two experiments, we investigated predictions of the *collaborative theory of language use* (Clark, 1996) as applied to instant messaging (IM). This theory describes how the presence and absence of different grounding constraints causes people to interact differently across different communicative media (Clark & Brennan, 1991). In Study 1, we document how IM changes as users increase in expertise. In Study 2, we compare adaptations across telephoning and IM with a focus on multitasking.

Keywords: grounding, instant messaging, multitasking, expertise, discourse markers, backchannels

Grounding in Instant Messaging

Instant messaging (IM) can be considered impoverished because it lacks information conveyed by the voice, face, and hands. But another way of looking at communication is as a system whose constraints have both positive and negative effects. The loss of visual and auditory information, for example, can be seen as a benefit if it allows briefer, more direct communication (e.g. no need to explain a disheveled appearance) in a setting that values brevity and directness. As people become more accustomed to a communicative technology, they adapt to the costs and benefits of that technology.

On the modern college campus, there are few students who have never used IM. For ease of exposition, we label the less experienced users *novices* and the more experienced users *experts*. In two studies, we tested nine grounding constraints. In the first study, participants of varying levels of expertise engaged in an IM conversation in a psycholinguistic laboratory. We report on adaptations observed from novice to expert instant messengers. In the second study, participants of varying levels of expertise answered a survey on multitasking. We report on the use of multitasking in instant messaging versus telephoning, and also on adaptations observed from novice to expert instant messengers.

Why IM Has the Characteristics It Has

There are two approaches to explaining why IM has the characteristics it has. One is that the characteristics developed idiosyncratically. The recognizable novel register may emerge from a combination of familiar registers, such as *note-taking register* and *postcardese* (Ferrara, Brunner, & Whittemore, 1991, pp. 12-13). In this approach, learning IM is a matter of learning a cultural more. The use of emotional expressions such as emoticons, capitalizations, and orthographic sounds such as *argh* are seen as forms of play (Cherny, 1999), rather than as

predictable products of the advantages and disadvantages of the communicative medium. Also, because people are thought to learn norms "haphazardly" (Kiesler, Zubrow, Moses, & Geller, 1985, p. 81), no predictable relationship between expertise and behavior is expected.

An alternative approach is that IM has the characteristics it has because expertise exerts predictable forces on the communicative system. This approach relies on the *collaborative theory of language use* (Clark, 1996). In this theory, communicators take into account each other's *common ground* in preparing and interpreting communication (Clark, 1996). Common ground includes things like prior history, shared environments, and information from the discourse at hand. As they take turns making contributions to the discourse, communicators check each other's understanding by *grounding* their contributions – making sure that what is contributed is understood well enough for current goals (Clark & Wilkes-Gibbs, 1986).

Theoretically, the way people ground will vary across different media, such as face-to-face dialogues versus written letters versus cell phone calls (Clark & Brennan, 1991). Because it shares elements with both written and spoken communication, the communicative medium of IM presents a unique setting in which to explore the predictions of the collaborative theory of language use. Much like earlier research has explored how grounding changes over time, such as across successive iterations of a referential card task (Clark & Wilkes-Gibbs, 1986), or across different communicative settings, such as monologues or dialogues (Fox Tree, 1999; Fox Tree & Mayer, 2008), or across different conversational participants, such as people familiar or unfamiliar with shared materials (Isaacs & Clark, 1987), we test how IM changes depending on communicators' experience and familiarity with the medium.

Grounding Constraints and Hypotheses

Predictions for the domain of instant messaging will be discussed in terms of nine

grounding constraints, seven discussed by Clark and Brennan (1991; visibility, audibility, cotemporality, simultaneity, sequentiality, reviewabilty, reviseability) and two new constraints we present here (multi-tasking and multi-tasking awareness). Like Clark and Brennan (1991), we do not want to imply that this list is exhaustive. There are many constraints that we do not test that would also have consequences for how language is produced and understood in IM, such as anonymity, synchronicity awareness, and locatability (Hård af Segerstad & Ljungstrand, 2002) or visual copresence, which is relevant for tasks involving references to objects that both communicators can see (Gergle, Millen, Kraut, & Fussell, 2004). Copresence is the eighth constraint discussed in Clark & Brennan (1991). We do not discuss it further because we did not have predictions for how lack of copresence (our experimental design) might be handled differently depending on expertise.

The *visibility constraint* is that communicators see each other while communicating, and the *audibility constraint* is that speech is used to communicate (Clark & Brennan, 1991). One benefit of visibility and audibility is that they allow communicators to use gestures, facial expressions, and prosody to interpret talk. As demonstrations of the outcome of these constraints, (1) gesture production overlapped less with words when there were visible addressees (Bavelas, Gerwing, Sutton, & Prevost, 2008), (2) communicators used hyperbole more in emails (71% of all nonliteral language; Whalen, Pexman, & Gill, 2009) than in talk (12% of ironic utterances; Gibbs, 2000) to make up for the lack of "paralinguistic cues" in emails (Whalen et al., 2009, p. 274), and (3) computer-mediated communicators used more verbal cues to convey emotion to make up for not being able to use nonverbal cues (Walther, Loh, & Granka, 2005). But visibility and audibility also have costs, if communicators must discuss something they could otherwise avoid discussing, such as a new scar or tremulous voice. As another example, ambient noise or

chewing affects audible communication but not written communication.

Hypothesis 1: Experienced IM'ers will be more likely to adopt orthographic stand-ins (e.g. emoticons, punctuation marks, phonetic spell-out of sounds) to compensate for the lack of intonational or gestural information.

In the current study, we do not differentiate between behaviors stemming from audibility versus visibility because the two are often linked. For example, in face-to-face communication, vocal emphasis can co-occur with other visible behaviors, such as raised eyebrows or expansive gestures. The capitalization in IM could be making up for the lack of visibility or audibility.

An alternative to this hypothesis is that the use of emotional expression will decrease with expertise. Although communicators produced more weighty emotional content in writing than in face-to-face communication (Kiesler et al., 1985), physiological arousal decreased over time (Kiesler et al., 1985). In addition, routinized computer behavior was less arousing than novel behavior (Kiesler et al., 1985). Although the researchers did not study expertise, a logical prediction based on these findings is that expertise engenders less emotional behavior. In support of this alternative, researchers have noted a lack of emoticons in the written communications of experienced writers (Ling & Baron, 2007; Whalen et al., 2009).

Both visibility and audibility have important consequences for the kind of information available in checking for addressees' understanding. *Negative evidence* of understanding is information that explicitly addresses a lack of understanding, such as saying *huh?* or displaying a quizzical expression. *Positive evidence* of understanding is information that confirms understanding has occurred, such as nodding or using backchannels such as *uh huh* while the other communicator holds the floor (see Clark & Brennan, 1991, for a discussion of positive and negative evidence). In IM, communicators must put evidence of understanding in writing, such

as by writing the negative evidence *Did you get my question? I asked what are bangers?* (example from current corpus) or by writing the positive evidence *I see*. IM'ers could possibly also use the timing of their messages to display lack of comprehension, such as by responding slowly, but this technique could be ambiguous because other factors, such as a momentarily failed internet connection, could also cause a slow message arrival time. Timing may be more informative in split-screen chat, where communicators can view each keystroke. Backchannels were noted to be distinctly lacking in some studies of split-screen communication (Kiesler, Siegel, & McGuire, 1984).

> **Hypothesis 2:** Experts will be more likely to understand the need for supplying positive evidence, and consequently will be more likely to transfer the *yeah*s and *uh huh*s of spontaneous talk to IM.

Novices will be more likely to treat IM like letter-writing, relying on negative evidence to indicated failed understanding. In support of this hypothesis, Cherny (1999, p. 197) found that experienced text-based communicators produced between 25 and 40 backchannels in 20 minutes.

The *cotemporality constraint* is that as communication is produced, it is understood, and the *simultaneity constraint* is that production and comprehension are possible at the same time (Clark & Brennan, 1991; this use of *simultaneity* differs from the use of the term to refer to multitasking, Cameron & Webster, 2005, p. 90). Face-to-face and telephone conversations are cotemporal and simultaneous. This allows speakers to alter planned talk mid-utterance based on addressees' reactions to words as they are being produced. Communication over walkie-talkie is cotemporal but not simultaneous; as speech is produced, it is understood, but only one person can speak at a time (the walkie-talkie is set in either *talk* or *listen* mode). This means that speakers cannot note reactions to what was said until the walkie-talkie turns have been turned

over to addressees.

IM is neither cotemporal nor simultaneous. The lack of cotemporality and simultaneity in IM has comprehension costs because communicators do not know, as they are preparing their communication, how the respondent will react until replies are received. This contrasts with split-screen chat. Split-screen chatters can see messages as they are being typed, making public (and potentially informative) the rate of keystrokes and all revisions. IM further contrasts from traditional letters, which are also neither cotemporal nor simultaneous, in that turns are relatively quick. This means that the cost of planning communication in the absence of simultaneous feedback is less in IM than in a letter. But IM has the additional factor of a waiting recipient (when it is used in a sustained conversation), which introduces delay costs (Clark & Brennan, 1991). If a wait is too long, recipients may wonder why. Indeed, respondents may take a second turn during a long delay, causing subsequent confusion about turn order.

Hypothesis 3: Experts will try to accommodate waiting addressees by producing more frequent and shorter turns.

Novices will not compensate for the lack of cotemporality and simultaneity, treating IM like letter or email exchanges with longer, less frequent turn packages.

An alternative to the theory that communicators shorten turns to increase grounding is that turns are shortened to "speed transmission" (Cherny, 1999, p. 147). More experienced IM'ers may type faster, and faster typers may have shorter turns. This alternative will be tested.

At first glance, the presence of a waiting addressee might additionally be predicted to affect IM'ers use of *um*. That is, IM'ers might be predicted to use *um* to indicate upcoming delay (Clark & Fox Tree, 2002; Smith & Clark, 1993). But the built-in delay of the IM message preparation phase obviates the need for *um* to indicate an ongoing delay in production. Such an

um would be expected in split-screen chat where lack of timeliness is readily apparent. But in IM, *um*s are predicted to be more similar to the deliberate *um*s found in movie dialogue rather than the *um*s that facilitate coordination in spontaneous conversation. For example, IM'ers may choose to place *um*s at the beginning of an answer to a question to suggest that the response required thought.

Hypothesis 4: Experts will be more likely to use *um*s, although the *um*s should not be used in the same places in IM as they would in talk or split-screen chat.

The *sequentiality constraint* is that turns go in sequence (Clark & Brennan, 1991). Sequentiality helps prevent confusion about what is responding to what. IM obeys the sequentiality constraint in some ways, because writers generally take turns sending messages. But it violates the sequentiality constraint in others, because people can mistime their responses such that they send a message on the heels of a prior message, at the same time that their addressee sends a message, resulting in the addressee's message appearing between the two planned-adjacent messages. In addition, people can respond to different topic streams, leading to nonsequential and potentially confusing messages.

Hypothesis 5: Experts will have an easier time avoiding misorderings and recovering intended orders when a misordering occurs.

Novices are predicted to expect sequentiality and become confused by the lack of sequentiality, resulting in the creation of more turns that are out of sequence. Past research supports the argument that turn-taking can cause grounding and understanding difficulty: Addressees performed better on completing a referential card task through IM when the writer included a cue to indicate that she had completed a turn (Hancock & Dunham, 2001), eliminating turn-taking confusion.

At the same time, there is existing data that at first glance belies this prediction. In a sample of workplace IM conversations, 41% of the conversations produced by heavy IM users contained misorderings compared to only 15% of the conversations produced by light IM users (Isaacs, Walendowski, Whittaker, Schiano, & Kamm, 2002). But these conversations averaged under five minutes in length. In addition, the data evaluated were whether the IM conversations contained any misordered turns, rather than how many misordered turns. Even experts may not be able to prevent some misorderings, and novices may be able to avoid misorderings by responding slowly; in fact, in the workplace IM study, light dyads had 28 seconds between turns compared to 22 seconds between heavy dyads (Isaacs et al., 2002). Intermittent messages between co-workers in an office may be too brief to uncover sequentiality effects predicted in a sustained IM conversation. Sequentiality effects may also be more likely to be found with narrative-style conversations. In the workplace IM study, 87% of conversations focused on work related topics rather than personal ones (Handel & Herbsleb, 2002; Isaacs et al., 2002); the conversations were also frequently interrupted without openings and closings (Isaacs et al., 2002).

Other evidence also belies this prediction at first glance. In an hour-long task-oriented study of dyadic and triadic messages sent by technologically-savvy writers, almost 80% of the samples contained at least one misordered turn (McCarthy, Wright, & Monk, 1992). Once again, the number of misordered turns was not assessed, although more were observed in triadic communication (McCarthy et al., 1992). In addition, this high rate of misordering may have resulted from the task (as McCarthy et al., 1992, note). Relevant to our predictions, McCarthy et al. (1992) proposed that shortening turns may reduce misorderings. If experts are more likely to shorten turns (as we predict), they may also be more likely to produce fewer misordered turns.

We test this proposal.

The *reviewability constraint* is that senders can review communication before it reaches addressees, and the *reviseability constraint* is that senders can privately revise messages before addressees encounter them (Clark & Brennan, 1991). Face-to-face and telephone talk are neither reviewable nor reviseable, but letter writing and IM, to some extent, are. One consequence is that written mistakes carry more negative baggage than spoken ones, making them more important to prevent and more costly to repair (Clark & Brennan, 1991). This is because of the assumption that writing can, and therefore should, be privately revised and have fewer faux pas than speaking. In addition, the capacity to cut and paste earlier segments into the current dialogue should cause people to be more careful in what they write in IM than what they say. However, much as people can forget that they are being recorded, IM'ers may choose to ignore the durability of their communication, although durability may play a more prominent role with more complex tasks with defined goals (Gergle et al., 2004; see also McCarthy & Monk, 1994, for discussion of the role of window size). If IM'ers choose to ignore durability, then they would be less concerned about errors. Indeed, Hård af Segerstad and Ljungstrand (2002) describe the internet communication "e-style" as having less formal spelling and syntax: it is "playful, dynamic and speech-like" (pp. 153, 162).

> **Hypothesis 6:** Novices will be more likely to use formal language, such as proper capitalization and punctuation. Experts will be more likely to use informal language, such as abbreviations and misspellings.

Novices are predicted to be more sensitive to the durability of IM, and so more likely to treat IM like letters, producing turns that require less revision than experts' turns. Experts are predicted to be more likely to write as they speak, and produce less crafted messages.

An alternative approach is that communicators use abbreviations and misspellings as part of a register designed to increase the speed of communication (Simon, 2006). That is, writers use *u r* instead of *you are* because they believe addressees recognize the register, which is motivated by efficiency (Simon, 2006, p. 373). This alternative will be tested.

In addition to formal and informal language differences, experts are predicted to be more likely to adopt *discourse markers* like *you know*, *well*, and *I dunno* in their writing. Although they are often thought of as indicative of informal, even sloppy, speech (Fox Tree, 2007), they do serve communicative functions in the environment of spontaneous talk (Fox Tree, 2000, 2006, 2007; Fox Tree & Schrock, 1999, 2002).

Hypothesis 7: Experts will use more discourse markers.

In addition to these seven constraints discussed by Clark and Brennan (1991), we propose two additional grounding constraints: the *multitasking ability constraint* (henceforth *multitasking constraint*) and the *multitasking awareness constraint*. Basic research into multitasking has been identified as an area of future study (Condon & Čech, 2001). We explore how the ability to multitask, and the awareness of the ability to multitask, constrain communication. That is, we are treating the multitasking constraint and the multitasking awareness constraint not as equivalent to the activity of multitasking, but as descriptors of situations in which communicators change behavior (such as how language is produced and understood) based on their presence or absence.

The multitasking constraint is that people can engage in multitasking. People can multitask while talking face-to-face, such as when a host is preparing dinner while entertaining a guest, or a driver and passenger are talking. Phone calls can involve multitasking, particularly if the phone is mobile. In contrast, traditional letter-writing involves uninterrupted time in a

stationary location. With IM, however, multitasking may take place during existing interruptions brought about by waiting for a communicator's message.

Hypothesis 8: Communicators will multitask more with IM than with telephone communication, and multitasking will increase with increasing expertise.

Because multitasking is predicted to be easier to hide with IM, people should be more likely to multitask while IM'ing than while talking on the phone.

The multitasking awareness constraint is that communicators are aware of each other's multitasking. Sometimes multitasking is sanctioned, such as when IM is used in a work environment (Isaacs et al., 2002); sometimes it isn't, such as when a participant expecting a fully engaged listener on the phone hears the tapping of a computer keyboard. The telephone pauses which might ordinarily have been attributed to thoughtfulness on the part of the listener may now be attributed to distraction, and demonstrations of understanding, such as saying "uh huh," may now be understood as invitations to continue speaking rather than as indicators of true comprehension.

Hypothesis 9: Communicators will be more aware of multitasking with telephone communication than with IM, and multitasking awareness will increase with increasing expertise.

Because timing is critical to telephone conversations (Clark, 2002), communicators should be acutely aware of multitasking while on the phone.

Study 1: Instant Messaging

Hypotheses 1 through 7 are tested. Predictions are that increasing expertise will lead to more orthographic stand-ins for emotion, more positive evidence of understanding (backchannels), shorter turns, more *um*s, fewer misordered turns, less formal language and more

informal language.

Method

Participants. Eighty-eight UCSC undergraduate students participated in partial fulfillment of course requirements. In addition to students from the broader pool, novices were targeted for participation without their knowledge of the criteria for participation. As part of their registering to use the experiment sign-up system, potential participants answered a variety of questions including demographic questions, handedness questions, and questions about whether they speak a second language. One of these questions was *How frequently do you use an instant messaging (IM) service*? Later, participants who claimed to use IM infrequently were contacted and asked to sign up for the current experiment, but they were not informed of the basis of the invitation. Participants' answers to other experiment sign-up questions were neither accessed nor recorded by us (they were used by other researchers in the department for other experiments).

Materials. A survey was designed to assess participants' experience with IM. The following questions were asked: (1) How frequently do you use an instant messaging (IM) service? with the following possible responses: *never*-1, *rarely*-2, *montly*-3, *weekly*-4, *daily*-5, (2) How comfortable are you with IM services? with answers ranging from *not comfortable at all*-1 to *very comfortable*-5, and (3) Rate how well you understand how to use an IM service with answers ranging from *complete novice*-1 to *expert*-5. A composite measure was created by averaging the three responses.

Procedure. Participants were tested in groups of two. Each was seated in front of a computer in individual rooms and given instructions to have a conversation with their partner using MSN Instant Messenger. Instructions indicated that participants should begin the conversation by exchanging stories either about a party that they had recently been to or a bad

roommate experience, but that they did not need to stay on topic. Participants exchanged IMs for about 20 minutes. Following the experiment, participants completed the survey assessing IM experience.

Coding. Three to five coders blind to the hypotheses (and without knowledge that IM expertise was evaluated at all) coded transcripts for the following (discrepancies in coding were resolved through discussion):

(1) ***Emotional expressions***. These included all orthographic stand-ins for emotions, such as emoticons, spelling words to imitate sounds (*ooh* and *aw*), laughter (*ha ha* and *lol* [laugh out loud]), putting a phrase in all capitals for emotional emphasis, using multiple exclamation points, putting words in quotes to indicate ironic stance, and swear words.

(2) ***Formal elements.*** Proper use of capitalization and punctuation.

(3) ***Informal elements.*** Misspellings and abbreviations, including shortened versions of words or phrases (such as *are you* written as *r u*) and contractions.

(4) ***Discourse markers.*** The complete list of items treated as discourse markers follows: *I mean*, *you know*, *well*, *I dunno*, *oh*, *ah*, *hm* (when not a backchannel), *like*, *and then*, *anyway so yeah*, *so then*, *oh well*, and *so anyway*. Defining discourse markers can be tricky (Fraser, 1999). We used the rule of thumb that the expression was not syntactically or semantically required by the sentence, coupled with unanimous rater agreement that the item was a discourse marker.

(5) ***Backchannels.*** Unlike discourse markers, backchannels were an unrestricted set. They were defined as responses made to a communicator's message that did not take the floor or interrupt the communicator. They never overlapped with discourse

markers. Some examples are *I see, sure*, *right*, *okay, weird,* and *you said it*.

(6) ***Yeahs.*** *Yeah*s included the spellings *yes*, *yeah, yea, ya,* and *yah*. Only *yeah*s that were not used in answer to a direct question were counted.

(7) ***Ums.*** Any spelling of *um* and *uh* were included in the category *ums*, such as *uhm*, *uhh*, and *umm*. There were two *er*s in the corpus, but they were produced by one IM'er and used exclusively for error correction. They were not included in the analysis.

In addition to these codings, one of the authors, blind to IM expertise, counted the turn misorderings in each dyad. Identifying misordered turns required skill at reading and interpreting IM transcripts. See Table 1 for an overview of the relationship between phenomena and constraints.

Results

Participants scored on average 3.57 (*SD* = 1.13, responses ran the full range of 1 to 5) on the composite scale of experience with IM. The median score was 4, underscoring the observation that most college students are comfortable with IM, know how to use it, and use it frequently. Based on the distribution of scores, we labeled participants whose scores fell below 4 as novices in comparison to those whose scores were 4 or more, whom we labeled experts. According to these definitions, the study had 15 novice-novice dyads, 13 novice-expert dyads, and 16 expert-expert dyads.

Responses more than two standard deviations from the mean were treated as outliers and removed from analyses. A Pearson's *r* was calculated between expertise and variables of interest. A *p* value of .006 was required for significance at the .05 level. This reflects a Bonferroni adjustment for the nine principal tests: One each for Hypotheses 1, 3, and 7, and two each for

Hypotheses 2, 5, and 6. Hypothesis 4 was not testable with a correlational analysis. Tests of competing hypotheses are not included in the adjustment. The actual *p* is reported in the results section below for significant results, but discussion is based on the Bonferroni correction.

Hypothesis 1 was marginally supported. People tended to use more emotional expressions the more experienced they were with IM, $r(84) = .24$, $p = .02$ (two outliers). The mean rate of emotional expressions was 2.3% (*SD* = 1.7%, exclusive of outliers).

Hypothesis 2 was supported. With more experience, participants used more backchannels, $r(85) = .39$, $p < .001$ (one outlier). The mean rate of backchannels was .6% (*SD* = .7%). The average number of backchannels in these 20 minute samples was 2.3 (*SD* = 3), which is far fewer than the 30 or so observed by Cherny (1999) in her samples, which may be attributed to different conversational topics, different conversational goals, or technological short-cuts (in Cherny's sample, a single keystroke could read as a backchannel). Also, with more experience, participants tended to use more *yeah*s, $r(84) = .28$, $p = .01$ (two outliers). The mean rate of *yeah* was 1.4% (*SD* = 1%, exclusive of outliers).

Hypothesis 3 was supported. People sent more messages the more experienced they were with IM, $r(85) = .44$, $p < .001$ (one outlier). The mean number of messages sent was 50 (*SD* = 25, exclusive of outlier). They also tended to create messages with fewer words per turn, $r(78) = -.28$, $p = .01$ (eight outliers). The mean number of words per message was 7.8 (*SD* = 3, exclusive of outliers). There was no support for the contrasting theory that communicators shorten turns to "speed transmission" rather than increase grounding (Cherny, 1999, p. 147): There was no relationship between IM experience and typing speed, $r(42) = .06$, $p = ns$ (given unknown comprehension and planning time, typing speed was assessed in terms of pairs' performances), and there was also no relationship between typing speed and turn length, $r(42) =$

-.14, $p = ns$. In addition, we found no evidence that communicators attempt to match turn length to their partners, $r(42) = .24$, $p = ns$ (cf. Condon & Čech, 2001).

Hypothesis 4 was partially supported. Because the corpus contained only 29 *um*s produced by 20 IM'ers, a correlational analysis was not possible. Part of the hypothesis was supported by a comparison of the uses of *um* in the IM corpus versus in spoken corpora. Almost every *um* in the corpus (93%) occurred either at the beginning of a turn or on its own turn (23) or near the beginning of a turn (as in *so uhm*; 4). The two exceptions were an *um* that occurred in a list and fell at the end of a turn and an *uh* that could possibly be seen as the beginning of a quotation (*and so its like..uh i cant walk THAT far*). This rate of turn-initial *um*s is almost twice that found in spontaneous spoken monologues (47%, Swerts, 1998, p. 490) and almost four times that found in spontaneous spoken dialogues (23%, Bortfeld, Leon, Bloom, Schober, & Brennan, 2001, p. 137).

In written-message-initial location, *um* is unlikely to be warning addressees of an upcoming delay, the function proposed for spontaneous speech (Clark & Fox Tree, 2002; Smith & Clark, 1993). Instead, *um* is used to imply that thought was needed in constructing a response, an inference that builds up from the delay-warning function (Fox Tree, 2002). Of the 29 *um*s in the IM corpus, 18 (62%) initialed a turn responding to a question. As further evidence, the 4 *um*s on their own line did not show a time-stamp pattern consistent with warning upcoming delays. In one, the current IM'er did indeed take a long time to send the message following the *um* (27 sec). But in another, the IM'er responded within one second after the *um*. In the remaining two, the addressee took a turn six seconds and eight seconds following the *um*.

Hypothesis 5 was marginally supported. A dyad expertise score was calculated by averaging the expertise scores of the two participants in a dyad. A rate of misordering was

calculated for each dyad by dividing the number of misorderings by the number of messages sent. Novice pairs tended to misorder more turns, $r(41) = -.28$, $p = .08$ (one dyad had outlying data; removing this dyad created a new outlier, the removal of which revealed a stronger relationship, $r(40) = -.36$, $p = .02$.). The tendency, as opposed to reliability, may reflect the fact that simultaneous turns happen even among experts. The mean rate of misorderings was 4.5% (*SD* = 4.5%), which is substantially less than that observed in workplace chat (Isaacs et al., 2002) or task-oriented communication (McCarthy et al., 1992). Misorderings were more common with longer turns, as McCarthy et al. (1992) also predicted, $r(42) = .30$, $p = .05$.

Hypothesis 6 was partially supported. Participants used less formal language the more experienced they were with IM, $r(81) = -.30$, $p = .006$ (five outliers). The mean rate of formal elements was 13% (*SD* = 7%, exclusive of outliers). But expertise did not affect the use of informal elements. There was no relationship between the rate of informal language and expertise, $r(84) = .11$, $p =$ ns (two outliers). The mean rate of informal language was 9.8% (*SD* = 3.5%, exclusive of outliers). In addition, there was no relationship between the rate of informal language and typing speed, $r(42) = .04$, $p = ns$, contradicting the proposal that abbreviations and misspellings result from a desire to speed communication (Simon, 2006),

Hypothesis 7 was supported. With more experience, participants used more discourse markers, $r(85) = .35$, $p = .001$ (one outlier). The mean rate of discourse markers use was 1.1% (*SD* = .8%).

Discussion

As people gain expertise in IM, they adopt methods designed to overcome the grounding constraints imposed by this communicative medium. The more expertise, the more likely the writer would be to use (1) orthographic stand-ins for emotions (tendency), (2) more frequent and

(3) shorter messages (tendency), (4) less formal language, (5) more discourse markers, (6) more backchannels, (7) more *yeah*s (tendency), and (8) the fewer misordered turns (tendency). Expertise had no relationship to the amount of informal language, as defined by misspellings and abbreviations.

There were too few *um*s in the corpus to conduct tests of the relationship between expertise and *um* production. However, an analysis of the *um*s available revealed that *um*s in IM differ from those found in spontaneously spoken monologues and dialogues. They were more likely to be used at the beginning of turns, rather than the middle or end, and to be used with particular implicatures. That is, instead of indicating an actual delay, as they might do in speaking or split-screen chat, they indicate that thought was needed in constructing a response.

Some phenomena identified in earlier studies of written communication, attributed to register formation, were absent in the current data, including *emoted verbs* (quotatives formed by replacing the quotative verb with a subsequent word, such as "Tom cools" instead of "Tom said, 'cool'," Cherny, 1999, p. 87) and certain "common" contractions such as *onna* for *on the* (Cherny, 1999, p. 87).

Study 2: Multitasking Survey

Hypotheses 8 and 9 were tested. We surveyed people's frequency of multitasking, whether they hid multitasking, and how they detected multitasking. We tested adaptations by expertise as well, but because the participants were not invited to participate based on their level of IM expertise, the expertise range was truncated. Little is known about what communicators' attitudes towards multitasking are, in either the telephone or the IM domain. In addition, little is known about deception and multitasking, including how often people try to hide the fact that they are multitasking and how people detect that someone is multitasking during a conversation.

Method

Participants. Fifty-six UCSC undergraduate students participated in partial fulfillment of course requirements.

Materials. A survey was designed to assess participants' experience with multitasking in IM and on the phone (telephoning was described as talking via either a land line or a cell phone). The following questions were asked (illustrated here with IM; the same questions were asked in another part of the survey using the word *phoning* instead of *IM'ing*):

(1) How often do you do something else (multitask) while IM'ing? (Estimate frequency between 0% and 100%)

(2) How often do you try to hide the fact that you are multitasking while IM'ing? (Estimate frequency between 0% and 100%)

(3) Have you ever caught someone else pretending not to multitask while IM'ing when they were, indeed, multitasking? What gave them away? (How did you know they were multitasking while IM'ing?)

Procedure. Half the participants received the questions about IM first, followed by the same questions for telephoning; the other half received questions about telephoning first, followed by IM. In addition to the multitasking survey, participants completed the survey of experience with IM described in Experiment 1.

Coding What Gave Away Multitasking. Responses for what gave away multitasking activity varied idiosyncratically. Comments on telephone indications of multitasking included "noise discrepancy," "detachment of person," "the sound of a keyboard," "distracted voice," "breathe weird," and "answering many questions or comments the exact same way." Comments on IM multitasking activity included "aloof behavior," "monosyllabic responses," "inadequate

answers," "a lot of 'uh huhs'," "forgot what we were talking about," and "talked to someone else they were talking to and was told what they are doing." In addition to this wide variety of responses, many participants commented about the timeliness of responses. Because timeliness was the only relatively common response, this open-ended question was content analyzed for whether or not timing was mentioned as the give-away clue.

Results

Because novice users were not targeted for this study, the average score on the composite scale of experience with IM was higher. Participants scored on average 4.24 (SD = .88, range of 1.33 to 5). The median score was 4.67.

In the following analyses, the degrees of freedom are not consistent because not all participants responded to all questions. A p value of .007 was required for significance at the .05 level (Bonferroni adjustment for seven tests). Paired sample t-tests were used to compare results for the IM and telephone surveys. Pearson's rs were used to test correlational hypotheses, such as the relationship between expertise and multi-tasking rates. An independent samples t-test was used to compare expertise scores across those who did and did not detect multitasking.

Hypotheses 8 was supported and Hypothesis 9 was marginally supported. Participants estimated their rate of multitasking in IM as higher than their rate of multitasking on the phone, 82% to 59%, paired $t(53) = 5.3$, $p < .001$. This estimated rate of multitasking with IM is similar to the percentage of workplace IM conversations in which at least one participant multitasked, 85.7% (Isaacs et al., 2002). But when they hid their multitasking, participants tended to estimate their rate of hiding as higher on the phone, 23% to 9%, paired $t(52) = 2.7$, $p = .009$. Correspondingly, 57% of participants reported hiding multitasking in phone calls, but 30% reported hiding multitasking in IM. The fact that participants multitask more in IM and that they

hide their multitasking less in IM supports the idea that multitasking in IM is more acceptable than multitasking in phone calls.

When multitasking activity was detected in an addressee, mis-timed responses were a frequent cause. Of the 36 participants who provided responses about what gave away multitasking while phoning, 22% mentioned timing. Of the 25 participants who provided responses about what gave away multitasking while instant messaging, 56% mentioned timing. Timing is critical in conversations (Clark, 2002), and it is also important in IM.

Although there were fewer novices in Study 2 than in Study 1, experience still modulated results. Notably, the more experienced participants were with IM, the more they multitasked in IM, $r(52) = .52$, $p < .001$. This finding is similar to those found in workplace IM where heavy IM users were found to multitask more (Isaacs et al., 2002). At the same time, the propensity to multitask was also a characteristic of an individual. Participants who multitasked more in one communicative medium also multitasked more in another, at least when comparing IM and phone behavior, $r(52) = .38$, $p = .005$.

Hiding multitasking did not follow this pattern, however. There was no evidence that the participants who hid multitasking in one domain also hid multitasking in another, $r(51) = .19$, $p = ns$. In addition, participant's levels of expertise with IM was unrelated to the percentage of time they hid multitasking in IM, $r(52) = .09$, $p = ns$. However, participants with more expertise with IM tended to detect others' IM multitasking more. Those who detected multitasking had an average expertise score of 4.49 ($SD = .66$; N = 25) compared to those who did not detect multitasking, whose average expertise score was 4.11 ($SD = .92$; N = 30; $t(53) = 1.7$, $p = .09$).

Discussion

Participants reported multitasking more with IM than when talking on the phone, and

hiding their multitasking more when on the phone. The most frequent source of detection of multitasking was timing. Because of the presence of a waiting addressee, timing was predicted to play a prominent role in both telephoning and IM. The current participants identified timing as a give-away clue more often with IM. Timing may be a stronger give-away for IM because of the effort required to produce a written backchannel message versus saying *uh-huh* while on the phone.

IM experience did influence IM multitasking. Those with more experience multitasked more, although they did not hide their multitasking more. Those with more experience also tended to detect multitasking more readily in others.

General Discussion

With the prevalence of instant messaging as a learning tool on college campuses, it is important to examine how people communicate with this medium. Grounding constraints change across media, and both students and teachers must compensate. For example, in a traditional classroom, students are able to demonstrate both positive and negative evidence of understanding with verbal and visual responses. When students use instant messaging, positive and negative evidence must be conveyed differently. Learners must recognize an effective way to demonstrate understanding, and teachers must be aware of the difficulty in doing so.

Communicators with more experience using IM will use it differently. They will write shorter and less formal messages. They will provide feedback about their understanding. And they will be less likely to create out of order turn sequences. Overall, experienced IM users will be more effective in handling the costs and benefits of each grounding constraint in instant messaging. It is important that instructors recognize these constraints and the challenges that

novice users may face, as it may be necessary for instructors to alter their communication accordingly.

In this paper, we looked at nine constraints on grounding and how they changed across varying levels of expertise. They were: visibility, audibility, cotemporality, simultaneity, sequentiality, reviewability, reviseability, multitasking, and multitasking awareness. The collaborative theory of language use posits that as people gain expertise, the way they handle the costs and benefits of each grounding constraint will change in predictable ways (Clark, 1996; Clark & Brennan, 1991). Results support predictions stemming from the theory. These results contrast with proposals that IM has the characteristics it has as a product of register formation.

One predictable set of changes revolve around the costs and benefits of the lack of visibility and audibility. With increasing experience, IM'ers tended to increasingly incorporate orthographic stand-ins for emotions ordinarily conveyed by facial cues or tone of voice. But one of the benefits of lacking visibility and audibility is the ability to engage in surreptitious multitasking. Correspondingly, we found evidence that people multitask more on IM than on the phone, possibly because the audibility of the phone allows multitasking to be detected. In addition, IM multitasking increased with increasing IM expertise. That is, as people became familiar with IM, they increasingly availed themselves of the opportunity to multitask.

Experience with a medium will change communicator's expectations. For example, on the phone, seconds between turns can influence inferences. In a study of overhearers' judgements of others' conversations, turn intervals that were two or three seconds longer led to increased ratings of production difficulty, dishonesty, and discomfort with a topic (Fox Tree, 2002). In IM, a few seconds between turns could result from something as mundane as typing speed. But timing can still be important in IM if the inter-turn wait does not match the message

length. For example, a written *uh-huh* that took thirty seconds to send might be suspect, but a turn that was three lines long and took thirty seconds to send might not be. In the current data, an IM'er sent a message with a single question mark after a sixty second delay awaiting the response to a question; the response arrived four seconds later and was fifty words long. As people become experienced with IM, they should develop a better sense of how long it takes to prepare messages of different lengths. In support of this, a trend was observed in the multitasking study that experience with IM was associated with better detection of multitasking in others.

As people gain experience with IM, they increasingly shift away from treating IM like writing and towards treating IM like speaking. Novices used more proper capitalization and punctuation, and experts used more discourse markers. However, it would go too far to say that uses in IM must mimic uses in talk. Most *um*s used in IM are at the beginning of a turn, but most *um*s used in talk are in the middle. Instead of indicating an actual delay, as in talk, *um*s in IM seem to indicate that the message required thought, whether or not there was an actual delay between typing *um* and typing the rest of the message.

Unlike the use of capitalization and proper punctuation, the use of misspellings and abbreviations did not vary by IM expertise. On the one hand, this supports the notion that IM is a casual communicative setting (Cameron & Webster, 2005) because misspellings are tolerated. One other hand, this conflicts with the prediction that informal language should increase with expertise. One possible resolution is that misspellings are rife in all writing of the population under study. Another is that the use of abbreviations may cross over from other communicative domains, such as text messaging (Herring, 2004).

Experience with IM does change the way people use the medium. In some sense, there is a change in the way people communicate. But this is not to say that IM'ers lose the ability to

hold a conversation. Instead, there are predictable ways that new communicative media can be expected to impact how communication proceeds. Because each medium has different benefits and costs, one medium should not necessarily replace another. Herring noted that by 2004, communication with streaming audio and video had not replaced text-based conversations, which she described as "CMC has yet to embrace its full multimedia potential" (Herring, 2004, p. 29). She argued that text-based communication may linger because it is "stable, simple, and usable across computing platforms" (Herring, 2004, p. 33). We demonstrate here that IM has unique advantages and disadvantages in comparison to other communication tools. Much as the telephone did not replace letter-writing, we believe increasingly multimodal forms of on-line communication (such as Skype) will not replace IM.

The collaborative theory of language use highlights the critical role that constraints on communication, purposes of communication, and means of communication have on the value of different forms of communication. Although IM can be seen as impoverished compared to face-to-face, it has unique costs and benefits. As a more dramatic example, Morse code is the best method to choose if the only means of communication is plumbing pipes and a rock. Less dramatically, companies are attempting to convert voice mail to text messages because people would rather read their messages than listen to them, even though reading strips communication of auditory information (http://marketplace.publicradio.org/display/web/2009/09/17/pm-voicemail/). The way people achieve grounding will change across media, and with experience, people will learn how to adapt to new media. We have demonstrated that here for IM. The same method can be applied to other media of the future.

References

Bavelas, J., Gerwing, J., Sutton, C., & Prevost, D. (2008). Gesturing on the telephone: Independent effects of dialogue and visibility. *Journal of Memory and Language, 58*(2), 495-520.

Bortfeld, H., Leon, S. D., Bloom, J. E., Schober, M. F., & Brennan, S. E. (2001). Disfluency rates in conversation: Effects of age, relationship, topic, role, and gender. *Language and Speech, 44*(2), 123-147.

Cameron, A. F. & Webster, J. (2005). Unintended consequence of emerging communication technologies: Instant messaging in the workplace. *Computers in Human Behavior, 21,* 85-103.

Cherny, L. (1999). *Conversation and Community: Chat in a Virtual World*. Stanford, CA: CSLI Publications.

Clark, H. H. (1996). *Using Language*. New York: Cambridge University Press.

Clark, H. H. (2002). Speaking in time. *Speech Communication, 36,* 5-13.

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, D. C.: American Psychological Association.

Clark, H. H., & Fox Tree, J. E. (2002) Using *uh* and *um* in spontaneous speaking. *Cognition, 84,* 73-111.

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

Condon, S. & Čech, C. (2001). Profiling turns in interaction: Discourse structure and function. In *Proceedings of the 34th Hawai'i International Conference on Systems Sciences, 4*. IEEE

Computer Society Press.

Ferrara, K., Brunner, H., & Whittemore, G. (1991). Interactive written discourse as an emergent register. *Written Communication, 8*(1), 8-34.

Fox Tree, J. E. (1999). Listening in on monologues and dialogues. *Discourse Processes, 27,* 35-53.

Fox Tree, J. E. (2000). Coordinating spontaneous talk. In L. R. Wheeldon (Ed.), *Aspects of Language Production* (pp. 375-406). Philadelphia: Psychology Press.

Fox Tree, J. E. (2002). Interpretations of pauses and *um*s at turn exchanges. *Discourse Processes, 34*(1), 37-55.

Fox Tree, J. E. (2006). Placing *like* in telling stories. *Discourse Studies, 8*(6), 749-770.

Fox Tree, J. E. (2007). Folk notions of *um* and *uh*, *like*, and *you know*. *Text & Talk, 27-3,* 297-314.

Fox Tree, J. E. & Mayer, S. A. (2008). Overhearing single and multiple perspectives. *Discourse Processes, 45,* 160-179.

Fox Tree, J. E., & Schrock, J. C. (1999). Discourse markers in spontaneous speech: Oh what a difference an oh makes. *Journal of Memory and Language, 40,* 280-295.

Fox Tree, J. E., & Schrock, J. C. (2002). Basic meanings of *you know* and *I mean*. *Journal of Pragmatics, 34,* 727-747.

Fraser, B. (1999). What are discourse markers? *Journal of Pragmatics, 31,* 931-952.

Gergle, D., Millen, D., Kraut, R.E., and Fussell, S.R. (2004). Persistence matters: Making the most of chat in tightly-coupled work. *Proceedings of the SIGCHI conference on Human Factors in Computing Systems,* pp. 431-438. New York: ACM Press.

Gibbs, R. W., Jr. (2000). Irony in talk among friends. *Metaphor and Symbol, 15,* 5-27.

Hancock, J.T., & Dunham, P.J. (2001). Language use in computer-mediated communication: The role of coordination devices, *Discourse Processes, 31*, 91-110.

Handel, M. & Herbsleb, J.D. (2002). What is Chat doing in the workplace? *Proceedings of the ACM conference on Computer Supported Cooperative Work*, pp. 1-10. New York: ACM Press.

Hård af Segerstad, Y., & Ljungstrand, P. (2002). IM'ing with WebWho. *International Journal of Human-Computer Studies, 56*(1), 147-171.

Herring, S. C. (2004). Slouching toward the ordinary: Current trends in computer-mediated communication. *New Media & Society, 6*(1), 26-36.

Isaacs, E. A., & Clark, H. H.. (1987). References in conversations between experts and novices. *Journal of Experimental Psychology, 116,* 26–37.

Isaacs, E., Walendowski, A., Whittaker, S., Schiano, D. J., & Kamm, C. (2002). The character, functions, and styles of instant messaging in the workplace. *Proceedings of the ACM conference on Computer Supported Cooperative Work*, pp. 11-20. New York: ACM Press.

Kiesler, S., Siegel, J., & McGuire, T. W. (1984). Social psychological aspects of computer-mediated communication. *American Psychologist, 39*(10), 1123-1134.

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

Ling, R. & Baron, N. S. (2007). Text messaging and IM: Linguistic comparison of American college data. *Journal of Language and Social Psychology, 26*(3), 291-298.

McCarthy, J. C. & Monk, A. F. (1994). Measuring the quality of computer-mediated

communication. *Behaviour & Information Technology, 13*(5), 311-319.

McCarthy, J. C., Wright, P. C., & Monk, A. F. (1992). Coherence in text-based electronic conferencing: coupling text and context. *Journal of Language and Social Psychology, 11*(4), 267–277.

Simon, A. F. (2006). Computer-mediated communication: Task performance and satisfaction. *The Journal of Social Psychology, 146*(3), 349-379.

Smith, V. L., & Clark, H.H. (1993). On the course of answering questions. *Journal of Memory & Language*, *32*, 25-38.

Swerts, M. (1998). Filled pauses are markers of discourse structure. *Journal of Pragmatics, 30,* 485-496.

Walther, J. B., Loh, T., & Granka, L. (2005). Let me count the ways: The interchange of verbal and nonverbal cues in computer-mediated and face-to-face affinity. *Journal of Language and Social Psychology, 24,* 36-65.

Whalen, J. M., Pexman, P. M., & Gill, A. J. (2009). "Should be fun -- not!" Incidence and marking of nonliteral language in e-mail. *Journal of Language and social Psychology, 28*(3), 263-280.

Table 1. Overview of Constraints and Coding for Study 1.

Constraint	Definition	Assessed By
Visibility and Audibility	Communicators can see and hear each other	• Orthographic stand-ins for emotional expressions
Cotemporality and Simultaneity	As communication is produced, it is understood; Production and comprehension can occur at the same time	• Number of words per turn • Number of messages sent • Positive evidence of understanding (backchannels, *yeah*s)
Sequentiality	Communicators' turns go in sequence	• Number of misordered turns
Reviewability and Reviseability	Communicators can privately review and revise communication before it reaches addressees	• Formal elements • Informal elements • Discourse Markers • Ums