

## Reconsidering Common Ground: Examining Clark's Contribution Theory in the OR

Timothy Koschmann  
Southern Illinois University  
[tkoschmann@siu.edu](mailto:tkoschmann@siu.edu)

Curtis D. LeBaron  
Brigham Young University  
[lebaron@byu.edu](mailto:lebaron@byu.edu)

**Abstract.** The constructs of "common ground" and "grounding" are frequently invoked in the CSCW literature as a mechanism by which participants engaged in joint activity coordinate their respective understandings of matters at hand. These constructs arise from a model of conversation developed by Herbert Clark and sometimes referred to as "contribution theory." We describe here the basic features of this theory and attempt to apply it in analyzing a fragment of enacted interaction. The interaction was recorded during an abdominal surgery performed with the aid of an endoscopic camera. We encountered difficulties, however, in applying contribution theory as an analytic framework within this concrete setting. We found further that the notion of common ground represents a confusing metaphor rather than a useful explanatory mechanism. We conclude with a suggestion that researchers in the future seek ways of constructing descriptions of joint activity that do not rely on the troublesome notions of grounding and common ground.

Such words insert a name in place of a problem, and let it go at that; they pull out no plums, and only say, "What a big boy am I!"

Dewey and Bentley (1991)

## Introduction

How collaborators make sense of their unfolding interaction has been a recurring topic within the CSCW literature, as well it should. Collaborators' meaning-making practices serve as the foundation upon which all cooperative work depends. The construct of "common ground" is frequently invoked as a mechanism by which joint understanding is achieved. McCarthy, Miles, and Monk (1991) described common ground as "the mutual knowledge, beliefs, and assumptions of the participants in a conversation" (p. 209). *Grounding*, then, is thought to be a process whereby "common ground is updated in an orderly way, by each participant trying to establish that the others have understood their utterances well enough for the current purposes" (p. 209).

The present paper reconsiders the notion of common ground and associated concepts. First, we review prior work in CSCW that builds on Clark's notion of common ground and revisit Herbert Clark's seminal writings on common ground (e.g., Clark & Brennan, 1991; Chark & Schaefer, 1989). We examine an actual fragment of interaction, an exchange that occurred in the operating room (OR) of a teaching hospital, and ask how does Clark's contribution theory and his notion of common ground help us explain what has transpired. We conclude by raising some concerns about employing grounding as an explanatory mechanism in CSCW research.

## Common Ground in CSCW Research

The effects of various media on grounding has been an important focus of research in CSCW. McCarthy et al., for example, compared two forms of textually mediated communication—pure on-line chat and on-line chat augmented with a shared "report space." Basing their predictions on Clark's contribution theory of discourse (Clark & Brennan, 1991; Chark & Schaefer, 1989), they hypothesized that the introduction of the shared report space window would facilitate grounding by constraining potential meanings through visibility and co-presence. In their study, pairs of subjects were asked to solve a design problem communicating only through the synchronous, text-based interface. For half of the dyads, the participants had a chat window and a private report space; the other half had chat windows and a shared report space. All subjects were asked to record their arguments and possible solutions. The groups were compared in terms of the number of solutions and arguments recorded in each pair by only one participant. This "degree of disagreement" measure was assumed to be inversely related to the development of common ground within the dyads.

A more recent experiment reported by Veinott, Olson, Olson, and Fu (1999) examined grounding in pairs of native English speakers (NS) and pairs of non-native English speakers (NNS) communicating either in an audio-only condition or an audio supplemented with 'talking-head' video. The researchers hypothesized that the introduction of video would more dramatically improve grounding for the non-native speakers who presumably start the conversation with less common ground. Pairs were asked to perform a joint map tracing exercise. The maps given to the two participants were similar, but not identical. Each speaking turn was coded using a pre-established set of categories. NS subjects had more turns coded as "giving instructions," defined as "first offering of new information to the partner" (p. 306), than the NNS pairs. Pairs were compared for time and accuracy of task completion. NS pairs performed better than NNS pairs, but NNS pairs showed more improvement through the introduction of video. Grounding was assumed to be directly related to task performance and the augmented communication medium was theorized to compensate for differences across groups of subjects in initial common ground.

Jackson, Anderson, McEwan, and Mullin (2000) reported on the effect of video frame rate on grounding. They had their subjects perform a variant on a "referential communication task" while communicating via a video channel. Two team sizes were employed: 2-member teams with each member working at different workstations and 4-member teams with two members working at each workstation. In addition to standard measures of communication and task performance, Jackson et al. examined the length of referential expressions. An established finding in psycholinguistics is that the length of referring expressions declines with repeated references to the same object. The experimenters found longer expressions of definite reference in pairs (but not foursomes) operating in the low-frame rate condition. In this case, the longer referring expressions can be interpreted as representative of greater effort toward grounding.

Fussell, Kraut, and Siegel (2000) also studied the effects of communicative medium on grounding. In their study, unskilled workers were paired with either an expert or novice helper and asked to perform manual repair tasks on a bicycle. A repeated measures design was used in which each dyad did one repair in each of three media conditions: (1) side-by-side, (2) partners separated but able to communicate via full-duplex audio, and (3) audio supplemented with video. In the video condition, the worker viewed a miniature monitor positioned directly in front of her/his right eye, while the helper was seated in front of a conventional desktop monitor. Both saw the same display consisting of pages from a repair manual along with two video windows. One video window displayed a head shot of the helper, while the other display captured video from a camera mounted on the worker's head. The experimenters hypothesized that "ease of conversational grounding, as indicated by message length, number of conversational turns, and use of deictic expressions, should be easiest in the side-by-

side condition and hardest in the audio condition" (p. 24). Their findings were consistent with this prediction and in agreement with Clark's contribution theory. Participant utterances were coded and special attention given to referential statements. They observed that deictic references were always possible in the side-by-side condition, but were only possible in the video condition when the worker's head-mounted camera was directed toward the object of reference. They also observed that in the video condition, only the worker could perform acts of ostensive reference, not the helper. This was consistent with the predicted finding that performance in the video condition approached that of the side-by-side condition only to "the extent to which the collaborators are able to use the video technology to facilitate grounding" (p. 24).

## Clark's Contribution Theory

The body of work summarized in the previous section orients to the question of how the media through which communication is conducted affect the ability of cooperating agents to accomplish a shared task. These affects are theorized in terms of differences in common ground and in the processes of grounding afforded by the media. The notion of common ground arises from a model of conversation advanced by Herbert Clark and his students. Clark's contribution theory extended the traditional sender/receiver model of communication by enlarging the frame of analysis from the single message unit (utterance) to an interactionally-developed "contribution." Clark and Schaefer (1989) described the model in this way:

In most traditional views, the speaker's job is to issue understandable utterances, and the listener's is to understand them. Conversations proceed utterance by utterance. In the collaborative view, the speaker and addressees try to do something more at the same time; establish the mutual belief that the addressees have understood what is uttered, to establish what the speaker meant as common ground. The process of contributing to a conversation consists of both specifying some content and grounding it, and the products are units we will call *contributions*. Conversations proceed, in this view, not utterance by utterance, but contribution by contribution. (p. 124)

Clark and Brennan (1991) stipulated that contributions to common ground have two phases:

*Presentation phase:* A presents utterance  $u$  for B to consider. He does so on the assumption that, if B gives evidence  $e$  or stronger, he can believe that she understands what he means by  $u$ .

*Acceptance phase:* B accepts utterance  $u$  by giving evidence  $e$  that she believes she understands what A means by  $u$ . She does so on the assumption that, once A registers that evidence, he will also believe that she understands. (p. 130)

After presentation, they described the recipient as being in one of four possible states:

State 0: B didn't notice that A uttered any *u*.

State 1: B noticed that A uttered some *u* (but wasn't in state 2).

State 2: B correctly heard *u* (but wasn't in state 3).

State 3: B understood what A meant by *u*.

(p. 130)

If B achieves State 3 with regard to *u*, then A is justified in placing it in common ground. Contributions to conversation, therefore, are contributions to common ground.

Clark and Brennan (1991) described how different media change the constraints on grounding in conversation. They list (p. 141) eight features of interaction: co-presence, visibility, audibility, co-temporality, simultaneity, sequentiality, reviewability, and revisability. They pointed out that face-to-face communication does not constrain interaction with regard to the first six features, but does not afford the last two. Electronic mail, on the other hand, offers *only* the last two features. Co-presence and visibility—the capability "to see and hear what each other is doing and looking at" and the capability to "see each other," respectively—are the most important for the discussion that follows.

In the corpus of directory assistance inquiries analyzed by Clark and Schaefer (1989), they found that conversational partners employed three methods of accomplishing acceptance: the partner presumes acceptance by making a new contribution, the partner asserts acceptance through some form of acknowledgment or continuer (e.g., *yes, uh hah, okay*), or the partner requests clarification of some or all of the contributor's presentation. The need for grounding may vary in different situations. Clark and Schaefer postulated that those engaged in conversation apply a *grounding criterion*, i.e. "the contributor and his or her partners mutually believe that the partners have understood what the contributor meant to a criterion sufficient for current purposes" (p. 129). This seems reasonable enough, as far as it goes, but it leaves open how such a criterion is established and satisfied in actual practice. To explore these matters in greater depth and to make this discussion a bit more concrete, we turn now to an instance of interaction in which a complex series of contributions are made.

## Contributing to Conversation in the OR

The setting within which we have chosen to study conversational "contributions" is the operating room (OR) of a busy teaching hospital. In the fragment analyzed here, there are three participants: one ("Attending") is a highly-experienced surgeon who reportedly has performed 1200-1300 surgeries of the type described here; another surgeon

("Resident") is in the final year of his residency, having participated in 80 to 90 of these surgeries; a third ("Clerk") is a medical student enrolled in a clerkship rotation. This operation was Clerk's first surgical experience. Attending, therefore, is providing guidance and supervision to Resident, while both Attending and Resident are responsible for giving instruction to Clerk.

The participants are performing a surgery known as a cholecystectomy—dissection and removal of the gallbladder. The gallbladder is a small, sack-like organ located near the liver. It receives and releases bile through the cystic duct and is supplied with blood by the cystic artery. To perform a cholecystectomy, therefore, a surgeon must isolate both the cystic duct and cystic artery, and then sever both vessels. This may sound straightforward enough, but the relevant body parts vary considerably from patient to patient. Furthermore, these body parts come swaddled in layers of connective and fatty tissue. So it is often difficult, even to the trained eye, to visually distinguish fine structures such as the cystic artery. Way finding within the abdominal cavity, consequently, is never a trivial matter and can challenge the skills of the most experienced surgeons.

Cholesysectomies are often performed endoscopically, a technique sometimes referred to as "keyhole surgery." Rather than make a single large incision, the surgeon inserts a fiber optic camera lens and special tools through small openings or "ports" in the patient's abdominal wall. Endoscopic surgeries interest us because of the manifold challenges to perception and coordination they pose to participants.<sup>1</sup> The use of video in endoscopic surgery differs from traditional video-mediated communication. The endoscopic camera enables co-located participants to jointly view a visual field not otherwise accessible to them—i.e., the interior of the patient's body. By revealing the potential referents of their conversation, the view afforded by the endoscopic camera represents an example of what Nardi et al. (1997) referred to as "video-as-data."

Here we analyze a videotaped fragment of interaction less than one minute in duration.<sup>2</sup> Our approach is microethnographic (see LeBaron, in press; Streeck & Mehus, in preparation). We account for human experience through careful descriptions of participants' vocal (Sacks, 1992) and visible (Kendon, 1990) behaviors, occurring naturally within unfolding strips of interaction, altogether situated within a social and material environment. Our emphasis on naturally occurring behaviors diverges from other research traditions that rely on invented examples, which propose how people might behave; surveys or journals, which depend on people's ability to recollect and account for their behavior; and data generated through experimental methods where

---

<sup>1</sup> See Koschmann, LeBaron, Goodwin, and Feltovich (2001) for further discussion of these challenges.

<sup>2</sup> A full transcript is provided in Appendix B. The transcription conventions, which are summarized in Appendix A, were developed by Gail Jefferson (see Atkinson & Heritage, 1984; Ochs, Schegloff, & Thompson, 1996).

subjects behave within laboratory conditions, removed from the people and things associated with their everyday lives. Microethnography adheres to principles of empirical social science: A particular phenomenon is taken to exist, to the extent that data, analyses, and conclusions are verifiable by others. Although participant observations, field notes, and interviews are acceptable data for microethnographic work, videotaped data have become a mainstay because they capture subtle details of interaction that analysts can review and others can verify. With ethnomethodological roots, microethnography privileges subjects' perspectives: Researchers avoid imposing their own theorized views on the social phenomena they examine by attending to the orientations and relevancies that the research subjects display (e.g., Potter, 1998).

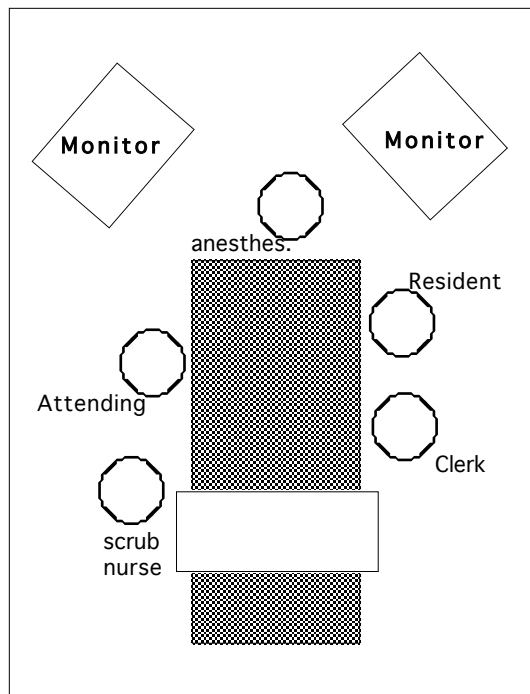
Microethnographic research has consistently shown that "human action is built through the simultaneous deployment of a range of quite different kinds of semiotic resources" (Goodwin, 2000, p. 2); that "talk in interaction shares billing with space, with artifacts, with work, and with the visible palpable body" (Moerman, 1990, p. 182). The physical arrangement of the operating room is shown in Figure 1. By maneuvering their bodies, participants oriented toward each other and the object of their collective work. Resident is located on the right side of the table and is primarily responsible for dissecting and removing the gallbladder. Attending assists (and supervises) from the left side of the table. Clerk stands adjacent to Resident and operates the rod lens of the endoscopic camera with close guidance from Resident. Looking within the endoscopic space is a team effort: Clerk steadies and directs the fiber-optic lens, Attending holds aside obstructing organs using a retractor tool, while Resident isolates relevant structure from surrounding connective tissue. One challenge is to correctly—and jointly—identify the cystic artery, as serious post-operative complications might arise if the wrong vessel were inadvertently ligated. The many ways in which the participants coordinate their activity, through understandings visibly displayed, seems beyond the reach of contribution theory, which is presently preoccupied with vocal pairs of presentation and acceptance.

As the transcribed moment begins, Attending (Lines 1 and 2) continues with a description of the surgical procedure for the benefit of Clerk, while Resident (Line 3) demonstrates the location of some abdominal feature.

Excerpt 1

- 1 Attending: Yeah (.) the other thing to do:: is make sure  
2                   you have your cystic (.) artery out too.  
3 Resident:                                        [(Right there)]  
4 Clerk:       Uh huh

Lines 1 and 2 appear as an example of “presentation” within a conversational contribution, as described by Clark and Schaefer (1989). Notice that Clerk's acceptance (line 4), the reduced continuer "Uh huh," is ambiguous in this situation because it is unclear whether it serves to accept Attending's statement concerning the cystic artery, Resident's demonstration, or both. Thus, we immediately encounter complexity not seen within the Clark and Schaefer report. Their examples consist exclusively of dyadic exchanges within telephone conversations. Here we have a moment of multi-party interaction with overlapping utterances in quick succession, making the relationship between presentations and acceptances ambiguous for participants and analysts alike.



*Figure 1: Organization of the workspace.*



Throughout our videotaped fragment, Resident is performing a blunt dissection to expose the cystic duct and the cystic artery using a tool in his right hand (a "black grasper"). His dissection is performed by burrowing the tip of the grasper into a bundle of connective tissue binding the bottom edge of the gallbladder to the common bile duct and then gently spreading apart the jaws of the instrument. He performs this continuously without performing any visible action that might be construed as a point. However, he occasionally provides a demonstrative reference, such as "Which is right back in here" (Line 5).

#### Excerpt 2

5 Resident: [Which is right back in here]  
6 Attending: [That way there is nuthin [ else before you  
7 h- (.) hit the edge of the liver  
8 (1.6)

Spatial deictics such as "here" were probably not common in the corpus examined by Clark and Schaefer, because the telephone participants were not co-present. Within the OR, however, spatial deixis is an integral part of the ongoing teaching work, as the experts orient themselves and their student within the uncharted spaces of this particular patient. Moreover, notice that Clerk does not respond to Resident's demonstrative reference at Line 5. In the conversational space in which Clerk might be expected to provide evidence of acceptance (Line 8), we have instead silence. What does silence mean? By remaining silent, does Clerk perform a kind of acceptance, showing Resident's demonstrative reference to be unproblematic and, therefore, not needing comment? Or does Clerk's silence represent a withholding of acceptance? To the extent that ambiguities of this sort are a pervasive feature of social life (and we think they are), Clark's contribution model is inadequate.

Ambiguities compound when Clerk eventually speaks at Line 10: "Can you see the cystic artery yet?"

#### Excerpt 3

9 Attending: That (kinda) guarantees you're safe too.  
10 Clerk: Can you see the cystic artery yet? I: 'd=

How should we categorize Clerk's query in terms of Clark's contribution theory? Who is Clerk addressing? In relation to Resident's prior talk (specifically the demonstrative reference at Line 5), Clerk's query may be regarded as an ongoing withholding of acceptance—that is, Clerk questions the location of what Resident says is "here." In relation to Attending's prior talk (lines 1, 2, 6, 7, 9), Clerk's query may be regarded as a presentation (i.e., moving the conversation forward through introduction of a new issue)

and hence an implied acceptance of what Attending has been saying. Within ongoing streams of multi-party interaction, the distinction between presentation and acceptance may be complicated as an utterance may connect forward or backward to various other utterances.

From Lines 11 to 25, Resident provides no less than five demonstrative references to the cystic artery's location.

Excerpt 4

11 Resident: =It's ↑r::ight back in the:re  
12 (2.1)  
13 Attending: °(We'll)° get it out here in a minute.  
14 (0.2)  
15 Resident: See it right there?  
16 (0.2)  
17 Clerk: U::mmmm=  
18 Resident: =Right (0.2) [there  
19 Clerk: [Okay yeah (.) yeah  
20 (0.2)  
21 Resident: That looks like the (0.2) where the money's a:t  
22 Clerk: Uhkay  
23 (0.2)  
24 Resident: En yih can see it's hanging out in that  
25 [tract.

Resident's behavior speaks to the environmental uncertainty of the current activity. The surgeons are literally moving through unfamiliar territory, changing (through dissection) the very scene that they are working to identify. This changing environment is at odds with Clark's contribution theory in that the theory would seem to require that contributions to common ground aggregate over time and remain relevant. The patient's abdominal cavity is not self-explicating: What objects are (and how they should be regarded) is something that the participants discuss and decide—not something they take for granted (see also Koschmann, et al., 2001). In this way, our data contrast with Clark and Brennan's (1991) discussion of "constraints" on grounding, which depicts the physical environment as fixed and readily accessible to co-present interlocutors. Our data shows how the component parts of the sender-receiver model (e.g., environment) may be brought into being through human interaction within and upon it.

When a physical environment is uncertain and unfolding, the relationship between presentation and acceptance may be muddled. After the first of Resident's five demonstrations, Clerk provides no response (Line 12). After the second demonstration, Clerk's response is delayed (Line 16) and non-committal (Line 17). Resident's third

demonstration is more than a vocal aside: He interrupts his dissecting work and uses the grasper tool as a prosthetic pointer in conjunction with his utterance (Line 18). It is then, and only then, that Clerk offers evidence of acceptance (Line 19). Nonetheless, Resident's demonstrations continue: He says "That looks like the (0.2) where the money's a:t" (Line 21); and "En yih can see it's hanging out in that tract" (Lines 24 and 25). With reference to Clark's contribution theory, the purpose of Resident's presentations in lines 21 and 24 is unclear. Since Clerk has already signaled acceptance at Line 19, why make two additional presentations? Resident's subsequent presentations would suggest a rejection of Clerk's acceptance, both in Line 19 and later in Line 22. Thus, our data show how the boundaries of a contribution may be ambiguous when conversation is tied to an ongoing activity.

In overlap with the end of Resident's last demonstration (Line 26), Attending makes a few critical presentations of his own.

#### Excerpt 5

24 Resident: En yih can see it's hanging out in that  
25                    [ɹ]tract.  
26 Attending: [ɹ](That's) actually big.°  
27 Attending: That's pretty bi:g.  
28 Attending: That may be ri:ght,  
29                    (0.4)  
30 Resident: That's right hepatic?

Attending says, "That's actually big" (Line 26); he repeats, "That's pretty big" (Line 27); and then he says, "That may be right" (Line 28). In this way, Attending invites more discussion of Resident's demonstrations, providing evidence that he does not accept Resident's presentation, raising the possibility that Resident has misidentified the cystic artery<sup>3</sup>. In turn, Resident provides evidence that he accepts Attending's non-acceptance (see Line 30). This exchange between Attending and Resident arguably constitutes a failure to achieve common ground, which highlights a gap in Clark and Schaefer's (1989) description: Although they describe methods by which participants in a conversation may achieve acceptance, they are silent regarding methods that contributors and their partners may employ to withhold acceptance. Additional questions remain. Does Clerk ever achieve state 3 with respect to identifying the cystic artery? Like Resident and Attending, we have no privileged access to Clerk's mental state at the time

---

<sup>3</sup> On the written page it might appear that the utterance "That may be right" (Line 28) is affirming. However, the prosodic cues and other features of Attending's delivery, along with his intent stare at the monitor, suggest another reading. Indeed, when he replied "That's right hepatic?" (Line 30), Resident's treated Attending's presentation as an incompletely formed utterance or question.

that this conversation occurred. We can only attempt to assess his understanding based on the accountable ways it is manifest within the observed interaction.

In the remainder of the fragment (Lines 31 through 41), Resident and Attending jointly decide to keep searching for the cystic artery.

Excerpt 6

31 (1.2)  
32 (Attending): [Comin' up]  
33 Resident: [The cystic may be up a little higher?=  
34 (Attending): =(Yup)  
35 Resident: You can see how easy that is to do we were  
36 just talking about that.  
37 Clerk: So you jus' dissect until you'r:e absolutely sure  
38 Attending: Ah [hah  
39 Resident: [Yeah=  
40 Clerk: =Till you see both the right hepatic and the cystic and then  
41 (4.6)

Resident treats the possible misidentification of the cystic artery as an instructable moment for the Clerk (Lines 35-36), who in turn formulates what may be learned from this exchange (Lines 37-40). Alternatively, if we chose to treat the series of demonstrative references (Line 11-30) as a side sequence, Clerk's extended utterance (Lines 37-40) could be seen as an acceptance of Attending's earlier description (Lines 1-9) of the surgical procedure. For participants and analysts alike, the relevancies of any utterance within an ongoing conversation are open to inference, question, and negotiation.

Did the Clerk ever learn to see and recognize the cystic artery? Although he participated in its surgical removal, and although he talked as one who understood (e.g., Lines 37-40), the Clerk's understanding remains an open question. Before the surgery, Attending asked Clerk to explain the "triangle of Calot" which is a region of the body that includes the cystic artery. The Clerk responded quickly and competently by naming abdominal features and their relative locations, including the cystic artery. However, he was evidently uncertain as to the location of these same anatomical features during the surgical procedure, suggesting that his textbook understanding did not translate into surgical practice. Moreover, we conducted interviews with Attending, Resident, and Clerk after the cholecystectomy. During our interviews, we played a video recording of their surgical procedure and asked each of them to identify (circle with a pen) the cystic artery when it appeared. The Clerk identified and circled something different than the others. According to our microethnographic study, the participants pursued shared understandings that had no obvious arrival. The tenuous nature of

mutual understanding within this practical setting contrasts with contribution theory's description of human interaction as an accumulation of contributions achieved.

By our reading, this sample of naturally occurring interaction poses certain challenges for Clark's theory. Commonplace ambiguities of everyday conversation (e.g., silence) frustrate applications of this theory to naturally occurring interaction. Within ongoing streams of interaction, the distinction between presentation and acceptance may be complicated as an utterance may connect forward or backward to various other utterances. Contributions, therefore, may seldom occur in tidy, recognizable packages. Complications associated with multi-party interaction (e.g., overlapping utterances) may make the relationship between presentations and acceptances ambiguous for participants and analysts alike. Contribution theory is, in the final analysis, a psycholinguistic account of a form of situated meaning making. Left out of this account are the embodied phenomena (e.g., gesture, gaze, facial expression, posture) that may play such an important role in human interaction. Also neglected are all the features of the material and social environment that participants draw upon in making sense of their own and others' utterances. These features are not static. When a physical environment is uncertain and unfolding, the relationship between presentation and acceptance may be muddled because relevancies for participants are constantly in flux.

## The Troubles with Common Ground

What is this thing called common ground? The research reports cited earlier from the CSCW literature would seem to suggest that it can be employed as an explanatory mechanism subject to experimental measurement, albeit indirectly. Our attempt to locate it in within naturally occurring discourse, however, would suggest that its status is considerably more ephemeral. It is at best a metaphor for a contingently-achieved and endlessly-defeasible state of alignment, independently inferred by each participant for every other participant to a conversation. The problem with the metaphor, however, is that, by its very nature, it tends to foster certain misconceptions about the phenomena it describes.

Clark (1996), in his more recent writings on common ground discusses it as a distributed form of mental representation. It is the superset of all of the sets of inferred understandings of *each* participant in a conversation independently maintained with respect to *every* other participant to the conversation. The superset is an abstraction, however, only observable by a god-like, omniscient outsider with privileged access to the participants' representational inventories. Serious problems arise when one begins to treat common ground as if it were a singularity, a possession of the participants, a place, an arrived at state, in short, as a noun instead of as a verb. McCarthy et al., (1991)

spoke of common ground as a singularity that is updated in an “orderly fashion.” The single entity, the superset, however, is only available to the omniscient outside observer, so what value does it hold for the participants themselves? Veinott et al. (1999) described achieving common ground as a negotiated process, but also spoke of it as possession of conversational partners (“when these pairs are provided with a video connection their performance improves to the point of being equal to those who *have* more basic common ground,” p. 308). Conversational partners who share a common language or cultural heritage would presumably have more inferred common understandings, but who possesses the common ground? Is it the individual participants? The pair? The conversation? In Jackson et al. (2000) the length of referring expressions is assumed to be a function of what has been placed in common ground. Fussell et al. (2000) studied “the ways in which the presence of visual information facilitates *grounding*” which they defined as “the development of mutual understanding between conversational participants.” In both cases, the researchers were attempting to measure changes in common ground as it was updated by the participants. Common ground cannot and should not be treated as an empirical fact. It is not a thing that can be measured, either directly or indirectly. Clark (1996), himself, makes this clear in more recent writings on the topic. There is a problem, however, with the notion of common ground even when it is understood in the more strict sense employed by Clark.

In prior writing (Koschmann et al., 2001) we have argued that Clark and Marshall’s (1981) model of reference repair dissects common ground in a way that obscures, but does not dispose of, the problem of intersubjectivity. Here we apply a similar critique to the notion of common ground itself. By its name it would seem to index a place, a place where things can be stored or recorded, but this is a profoundly misleading connotation. Common ground is, after all, a place with no place. It is a cooperatively constructed mental abstraction, available to no one. To paraphrase Dewey and Bentley, it inserts a name where the problem should be. Left out is an account of how participants in conversation routinely and unproblematically coordinate their understandings of matters at hand. Suggesting that they do so by placing propositions in common ground would seem to offer little in terms of conceptual understanding or descriptive power. What it provides instead is fertile ground for confusion. Perhaps it is time that we set this terminology aside and seek new ways to describe how people come to understand each other.

## Acknowledgements

We thank Herb Clark and four anonymous reviewers for numerous helpful suggestions. We also thank our collaborators on this project, Paul Feltovich, Chuck Goodwin, and Alan Zemel. Support came from a grant from the National Science Foundation.

## References

- Atkinson, J.M. and Heritage, J. (1984). Transcription Notation. In J.M. Atkinson & J. Heritage (eds.), *Structures of Social Action*. Cambridge University Press, New York.
- Dewey, J. and Bentley, A. (1991). Knowing and the Known. In J. A. Boydston (ed.), *John Dewey: The Later Works, 1949–1952, Vol. 16*. SIU Press, Carbondale, IL.
- Clark, H. (1996). *Using Language*. NY: Cambridge University Press.
- Clark, H. and Brennan, S. (1991). Grounding in Communication. In L.B. Resnick, J.M. Levine, and S.D. Teasley (eds.), *Perspectives on Socially-Shared Cognition*. APA, Washington, D.C.
- Clark, H. & Marshall, C. (1981). Definite reference and mutual knowledge. In A.K. Joshi, B.L. Webber, and I.A. Sag (Eds.), *Elements of discourse understanding* (pp. 10-63). NY: Cambridge University Press.
- Clark, H. & Schaefer, E. (1989). Collaborating on Contributions to Conversations. In R. Dietrich and C.F. Graumann (eds.), *Language Processing in Social Context*, Elsevier Science Publishers, North-Holland.
- Fussell, S., Kraut, R., & Siegel, J. (2000). Coordination of Communication: Effects of Shared Visual Context on Collaborative Work. In *Proceedings of CSCW 2000*, pp. 21-30.
- Goodwin, C. (2000). Action and embodiment within situated human interaction. *Journal of Pragmatics*, 32, 1489-1522.
- Jackson, M., Anderson, A., McEwan, R., & Mullin, J. (2000). Impact of Video Frame Rate on Communicative Behavior in Two and Four Party Groups. In *Proceedings of CSCW 2000*, pp. 11-20.
- Kendon, A. (1990). *Conducting interaction: Patterns of behavior in focused interaction*. Cambridge: Cambridge University Press.
- Koschmann, T., LeBaron, C., Goodwin, C., & Feltovich, P. (2001, August). Dissecting common ground: Examining an instance of reference repair. In J. D. Moore & K. Stenning (Eds.), *Proceedings of the Twenty-Third Annual Conference of the Cognitive Science Society* (pp. 516–521). Mahwah, NJ: Lawrence Erlbaum Associates

- LeBaron, C. (in press). Considering the social and material surround: Toward microethnographic understandings of nonverbal behavior. In V. Manusov (Ed.), *The Sourcebook of Nonverbal Measures: Going Beyond Words*. Mahwah, NJ: Erlbaum.
- McCarthy, J., Miles, V., & Monk, A. (1991). An Experimental Study of Common Ground in Text-Based Communication. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*, pp. 209-215.
- Moerman, M. (1990). Exploring talk and interaction. *Research on Language and Social Interaction*, 24, 173-187.
- Nardi, B., Kuchinsky, A., Whittaker, S., Leichner, R., & Schwarz, H. (1997). Video-as-Data: Technical and Social Aspects of a Collaborative Multimedia Application. In K. Finn, A. Sellen, S. Wilbur (eds.), *Video-Mediated Communication*, Lawrence Erlbaum, Mahwah, NJ.
- Potter, J. (1998). Cognition as context (Whose cognition?). *Research on Language and Social Interaction*, 31, 29-44.
- Sacks, H. (1992). *Lectures on conversation*. Oxford: Blackwell.
- Streeck, J., & Mehus, S. (in prep). Microethnography: The study of practices. In K. Fitch & R. Sanders (Eds.), *Handbook of Language and Social Interaction*. Mahwah, NJ: Erlbaum.
- Veinott, E., Olson, J., Olson, G., & Fu, X. (1999). Video Helps Remote Work: Speakers Who Need to Negotiate Common Ground Benefit from Seeing Each Other. In *Proceedings of CHI '99*, pp. 302-309.



## Appendix A: Transcription conventions

<b>Timing</b>		
Brackets	[ ]	Marks the beginning and end of temporal overlap among utterances produced by two or more speakers.
Equal sign	=	Indicates the end and beginning of two sequential 'latched' utterances that continue without an intervening gap. In some cases, the symbol is used in combination with brackets.
Timed silence	(1.8)	Measured in seconds, a number enclosed in parentheses represents intervals of silence occurring within (i.e. pauses) and between (i.e. gaps) speakers' turns at talk.
Micropause	(.)	A timed pause of less than 0.2 sec.
<b>Delivery</b>		
Comma	,	Indicates a continuing intonation with slight upward or downward contour that may or may not occur at the end of a turn constructional unit (TCU) as in the enunciation of an item in a not yet completed list.
Period	.	Indicates a falling pitch or intonational contour at the conclusion of a TCU.
Question mark	?	Rising vocal pitch or intonational contour at the conclusion of a TCU. An inverted mark represents a half rise.
Exclamation point	!	Marks the conclusion of a TCU delivered with emphatic and animated tone.
Hyphen	-	An abrupt (glottal) halt occurring within or at the conclusion of a TCU.
Colon(s)	:	A colon indicates sustained enunciation of a syllable vowel, or consonant. Longer enunciation can be marked using two or more colons.
Greater than/ Less than signs	> < < >	Portions of an utterance delivered at a noticeably quicker (> <) or slower (< >) pace than surrounding talk.
Degree signs	° °	Marks speech produced softly or at a lower volume than surrounding talk.
Capitalization		Represents speech delivered more loudly than surrounding talk.
Underscored text		Underscoring indicates stress on a word, syllable or sound.
Arrows	↑ □	Marks a rise or fall in intonation.
<b>Other</b>		
	hhh	Audible expulsion of breath (linguistic aspiration) as in laughter, sighing, etc. When aspiration occurs within a word, it is set off with parentheses.
	•hh	Audible inhalation is marked with a preceding dot.
Parentheses	( )	Text enclosed in parentheses represents transcribed talk for which doubt exists. Empty parentheses represent untranscribed talk or unknown speaker.
Double parentheses	(( ))	Transcript annotations (text italicized).

Appendix B: "Can you see the cystic artery yet?"  
Tape: #99-001B (0:08:48:10 to 0:09:34:55)

1 Attending: Yeah (.) the other [thing to do:: is make sure  
2 you have your cystic (.) artery out too.  
3 Resident: [Right there)  
4 Clerk: Uh huh  
5 Resident: [Which is right back in here]  
6 Attending: [That way there is nuthin [else before you  
7 h- (.) hit the edge of the liver  
8 (1.6)  
9 Attending: That (kinda) guarantees you're safe too.  
10 Clerk: Can you see the cystic artery yet? I:'d=  
11 Resident: =It's ↑r::ight back in the:re  
12 (2.1)  
13 Attending: °(We'll)° get it out here in a minute.  
14 (0.2)  
15 Resident: See it right there?  
16 (0.2)  
17 Clerk: U::mmmmmm=  
18 Resident: =Right (0.2) [there  
19 Clerk: [Okay yeah (.) yeah  
20 (0.2)  
21 Resident: That looks like the (0.2) where the money's a:t  
22 Clerk: Uhkay  
23 (0.2)  
24 Resident: En yih can see it's hanging out in that  
25 [tract.  
26 Attending: [That's) actually big.°  
27 Attending: That's pretty bi:g.  
28 Attending: That may be ri:ght,  
29 (0.4)  
30 Resident: That's right hepatic?  
31 (1.2)  
32 (Attending): [Comin' up)  
33 Resident: [The cystic may be up a little higher?=  
34 (Attending): =(Yup)  
35 Resident: You can see how easy that is to do we were  
36 just talking about that.  
37 Clerk: So you jus' dissect until you'r:e absolutely sure  
38 Attending: Ah [hah  
39 Resident: [Yeah=  
40 Clerk: =Till you see both the right hepatic and the cystic and then  
41 (4.6)