

Interaction Analysis of Dual-Interaction CSCL Environments

Murat Perit Cakir, Gerry Stahl, Drexel University, Philadelphia, USA
Email: MPC48@drexel.edu, Gerry@GerryStahl.net

Abstract: In order to collaborate effectively in group discourse on a topic like mathematical patterns, group participants must organize their activities so that they have a shared understanding of the significance of their utterances, inscriptions and behaviors—adequate for sustaining productive interaction. The need for participants to coordinate their actions becomes particularly salient in dual-interaction environments, where, e.g., chat postings and graphical drawings must work together; analysts of such interactions must identify the subtle and complex ways in which meaning making proceeds. This paper considers the methodological requirements on analyzing interaction in dual-interaction environments by reviewing several exemplary CSCL studies. It reflects on the nature of social organization, grounding and indexicality that frame the interaction to be analyzed.

“Everything that can be counted does not necessarily count;
everything that counts cannot necessarily be counted.” Albert Einstein

The Problem of Social Organization in Dual-Interaction Collaboration Spaces

A central issue in the theory of collaborative learning is how students can solve problems, build knowledge, accomplish educational tasks and achieve other cognitive accomplishments together. How do they share ideas and talk about the same things? How do they know that they are talking about, thinking about, understanding and working on things in the same way? Within CSCL, this has been referred to as the problem of the “attempt to construct and maintain a shared conception of a problem” (Roschelle & Teasley, 1995), “building common ground” (Baker et al., 1999; Clark & Brennan, 1991) or “the practices of meaning making” (Koschmann, 2002). We have been interested in this issue for some time: (Stahl, 2006) documents a decade of background to the Virtual Math Teams (VMT) research reported here—chapter 10 (written in 2001) argued the need for a new approach and chapter 17 (written in 2002) proposed the current VMT project. During the past six years (see Stahl, 2009), we have been studying how students in a synchronous collaborative online environment organize their interaction so as to achieve intersubjectivity and shared cognitive accomplishments in the domain of school mathematics.

Knowledge building in CSCL has traditionally been supported primarily with asynchronous technologies (Scardamalia & Bereiter, 1996). Within appropriate educational cultures, this can be effective for long-term development of ideas by learning communities. However, in small groups and in many classrooms, asynchronous media encourage exchange of individual opinions more than co-construction of progressive trains of joint thought. We have found informally that synchronous interaction can more effectively promote what we term “group cognition”—the accomplishment of “higher order” cognitive tasks through the coordination of contributions by individuals within the discourse of a small group.

In CSCL settings, interaction is mediated by a computer environment. Students working in such a setting must enact or invent ways of coordinating their understandings by means of the technological affordances that they find at hand. The development and deployment of these methods is not usually an explicit, rational process that is easily articulated by either the participants or analysts. It takes place tacitly, unnoticed, taken-for-granted. In order to make it more visible to us as analysts, we have developed an environment that makes the coordination of interaction more salient and captures a complete record of the group interaction for detailed analysis (Stahl, 2009, Ch 15 & 16). In trying to support online math problem solving by small groups, we have found it important to provide media for both linguistic and graphical expression. This resulted in what is known within CSCL as a *dual-interaction space*. In our environment, students coordinate their text chat postings with whiteboard drawings. A careful analysis of how they do this reveals as well their more general methods of social organization.

In this paper, we review approaches to dual-interaction spaces by important recent CSCL studies. The analytic thrust of these studies is to arrive at quantitative results through statistical comparisons of aggregated data. To accomplish this, they generally have to restrict student actions in order to control variables in their studies and to facilitate the coding of student utterances within a fixed ontology. We fear that this unduly restricts the interaction, which must be flexible enough to allow students to invent unanticipated behaviors (Dillenbourg & Jermann, 2006). The restrictions of laboratory settings make problematic experimental validity and generalization of results to real-world contexts. Even more seriously, the aggregation of data—grouping utterances by types or codes rather than maintaining their sequentiality—ignores the complexity of the relations among the utterances and actions. According to our analysis (Stahl, 2009, Ch 26), the temporal and semiotic

relations are essential to understanding, sharing and coordinating meaning, problem solving and cognition. While quantitative approaches can be effective in testing model-based hypotheses, they seem less appropriate both for exploring the problem of interactional organization and for investigating interactional methods, which we feel are central to CSCL theory. The effort to operationalize and quantify interactional phenomena risks reducing and reifying rich relationships into categories that fail to capture the group practices (Stahl, 2006, Ch 10). In particular, if individual chat postings are coded, then the unit of analysis becomes the individual action rather than the group interaction. Even if relations of neighboring actions are considered the often complex effects of temporally distant references upon the group meaning making is systematically ignored.

In the following, we will review studies of dual-interaction spaces in the CSCL literature in terms of their methodological orientation, underlying theoretical background and software features. The goal of our review is to offer a methodological and theoretical refinement of the conceptual frameworks underlying these studies. In the light of the common themes we identify across these studies we will argue that we need to conduct systematic case studies exploring the ways participants organize their interaction across multimodal interaction spaces in order to see how groups work on more open-ended tasks in less restricted online environments. Finally, we will propose a conceptual framework based on insights from ethnomethodology and linguistic anthropology that better suits the analytical challenges posed by the sequentially emergent and indexical nature of interactions mediated by CSCL environments with multimodal interaction spaces.

Approaches in CSCL to Analyzing Multimodal Interaction

Multimodal interaction spaces—which typically bring together two or more synchronous online communication technologies such as text chat and a shared graphical workspace—have been widely used to support collaborative learning activities of small groups (Dillenbourg & Traum, 2006; Jermann, 2002; Mühlfordt & Wessner, 2005; Soller & Lesgold, 2003; Suthers et al., 2001). The way such systems are designed as a juxtaposition of several technologically independent online communication tools carries important interactional consequences for the users. Engaging in forms of joint activity in such online environments requires group members to use the technological features available to them in methodical ways to make their actions across multiple spaces intelligible to each other and to sustain their joint problem-solving work.

In this paper, we review existing studies in the CSCL research literature that focus on the interactions mediated by systems with multimodal interaction spaces to support collaborative work online. We have selected sophisticated analyses, which go well beyond the standard coding-and-counting genre of CSCL quantitative studies, in which utterances are sorted according to a fixed coding scheme and then statistics are derived from the count of utterances in each category. Our review is not meant to be exhaustive, but representative of the more advanced analytical approaches employed. Unlike the simple coding-and-counting studies, the approaches we review attempt to analyze some of the structure of the semantic and temporal relationships among chat utterances and workspace inscriptions in an effort to get at the fabric of common ground in dual-interaction online environments.

The communicative processes mediated by multimodal interaction spaces have attracted increasing analytical interest in the CSCL community. A workshop held at CSCL 2005 specifically highlighted the need for more systematic ways to investigate the unique affordances of such online environments (Dillenbourg, 2005). Previous CSCL studies that focus on the interactions mediated by systems with two or more interaction spaces can be broadly categorized under: (1) prescriptive approaches based on models of interaction and (2) descriptive approaches based on content analysis of user actions.

(1) The *modeling approach* builds on the content-coding approach by devising models of categorized user actions performed across multimodal interaction spaces. We look at two examples:

- a. Soller & Lesgold's (2003) use of Hidden Markov Models and
- b. Avouris et al.'s (2003) Object-oriented Collaboration Analysis Framework.

In these studies the online environment is tailored to a specific problem-solving situation so that researchers can partially automate the coding process by narrowing the possibilities for user actions to a well-defined set of categories. The specificity of the problem-solving situation also allows researchers to produce models of idealized solution cases. Such ideal cases are then used as a baseline to make automated assessments of group work and learning outcomes.

(2) The *descriptive approach* informed by content analysis also involves categorization of user actions mediated by multimodal interaction spaces, applying a theoretically informed coding scheme (Neuendorf, 2002). Categorized interaction logs are then subjected to statistical analysis to investigate various aspects of collaborative work such as:

- a. The correlation between planning moves performed in chat and the success of subsequent manipulations performed in a shared workspace (Jermann, 2002; Jermann & Dillenbourg, 2005),
- b. The relationship between grounding and problem-solving processes across multiple interaction spaces (Dillenbourg & Traum, 2006),
- c. A similar approach based on cultural-historical activity theory (Baker et al., 1999), and

- d. The referential uses of graphical representations in a shared workspace in the absence of explicit gestural deixis (Suthers, Girardeau & Hundhausen, 2003).

(a) Soller and Lesgold's modeling approach involves the use of Hidden Markov Models (HMM) to automatically detect episodes of effective knowledge sharing (Soller & Lesgold, 2003) and knowledge breakdowns (Soller, 2004). The authors consider a programming task where triads are asked to use object-oriented modeling tools to represent relationships among well-defined entities. The task follows a jigsaw design where each group member receives training about a different aspect of the shared task before meeting with other members. The group sessions are hosted in the Epsilon online environment, which includes a text-chat area and a shared workspace. The workspace provides basic shapes that allow users to diagrammatically represent entities and relationships. Participants are required to select a sentence opener to categorize their contributions before posting them in the chat window. The authors manually extract segments from their corpus where each member gets the opportunity to share the unique knowledge element he/she was trained in with other group members. Some of these episodes are qualitatively identified as ideal cases that exemplify either an instance of effective knowledge sharing or a knowledge breakdown, completely based on the results of post-tests. For instance, a segment is considered an effective knowledge-sharing episode provided a chance for demonstrating the unique knowledge element comes during the session, the presenter correctly answers the corresponding questions in both pre- and post-tests, and the explanation leads at least one other member to correctly answer the corresponding question(s) in the post-test. The sequence of categorized actions (including chat postings and workspace actions) that correspond to these ideal cases is used to train two separate HMMs for the breakdown and effective knowledge sharing cases, respectively. An HMM computes the probability of a certain kind of action immediately following another; it thus captures certain aspects of sequentiality. These models are then used to automatically classify the remaining episodes and to assess team performance. However, the method is seriously limited to recognizing connections among actions to those based on immediate sequences of codes. While this can capture adjacency pairs that are important to conversation, it misses more distant responses, interrupted adjacency pairs, temporal markings and semantic indexes. The authors apparently make no specific distinction between workspace and chat actions as they build their HMMs over a sequence of interface actions. Moreover, the relationship between object diagrams constructed in the workspace and the explanations given in chat do not seem to be considered as part of the analysis. Hence, it is not clear from the study how a successful knowledge-sharing episode is achieved in interaction and whether the way participants put the affordances of both interaction spaces into use as they explain the materials to each other have had any specific influence on that outcome. Although they were reported to be successful in classifying manually segmented episodes, HMMs computed over a sequence of categorized actions seem to obscure these interactional aspects of the coordination of chat and workspace.

(b) The modeling approach outlined in Avouris et al. (2003) and Komis et al. (2002) proposes a methodology called the Object-oriented Collaboration Analysis Framework (OCAF) that focuses on capturing the patterns in the sequence of categorized actions through which dyads co-produced objects in a shared task space. The collaborative tasks the authors used in their online study included the construction of database diagrams with well-defined ontological elements such as entities, relationships and attributes. In this problem-solving context the final representation co-constructed in the shared workspace counted as the group's final solution. The OCAF model aims to capture the historical evolution of the group's solution by keeping track of who contributed and/or modified its constituent elements during the course of an entire chat session. The authors not only consider direct manipulation acts on specific elements but also chat statements through which actors propose additions/modifications to the shared diagram or agree/disagree with a prior action. The chat and drawing actions are categorized in terms of their functional roles (e.g., agree, propose, insert, modify, etc.). The mathematical model includes the sequence of categorized actions and the associations among them. The model is then used to gather structural properties of interactions (e.g., how contributions are distributed among dyads, what functional role each contribution plays) and to trace how each action performed in the interface is related to other actions. This modeling approach differs from similar approaches in terms of its specific focus on the objects co-constructed in the shared workspace. The model captures the sequential development of the shared object by keeping track of the temporal order of contributions made by each user. However, it is not clear from the study how the model could deal with the flexibility of referential work. For instance a chat posting may refer to multiple prior postings or to a sub-component of a more complicated entity-relationship diagram by treating several elemental objects as a single object. In other words, a model trying to capture all possible associations between individual actions in a bottom-up fashion may miss the flexibility of referential work and obscure the interactional organization.

(c) Jermann (2002) employs a coding scheme to study the correlation between planning moves in the chat area and the success of subsequent manipulations performed on the shared simulation in the Traffic Simulator environment. The shared task involved students tuning red-green periods of four traffic lights in the simulation to figure out an optimal configuration to minimize the waiting time of cars at intersections. The workspace could be manipulated in specific ways by users. The workspace also includes a dynamic graph that

shows the mean waiting time for the cars. The goal of the task is to keep the mean value below a certain level for two minutes. The study included additional experimental cases where dynamically updated bar charts are displayed to provide feedback to users about their level of participation. The logs of recorded sessions are coded in terms of their planning and regulatory content. The nature of the task allowed authors to numerically characterize different types of work organizations in terms of the distribution of manipulations performed on four possible traffic lights. The authors complement this characterization with number of messages posted, number of manipulations done and the types of messages as captured in the coding scheme. The study reported that dyads who coordinated their actions across both interaction spaces by planning what to do next (i.e., task regulation) and discussing who should do what (i.e., interaction regulation) in chat before manipulating the simulation performed better (i.e., achieved the objective more quickly). The interaction meters were not reported to have significant effects on promoting task and interaction regulation. The work of high performance groups are characterized with phrases like “posted more messages,” “more frequent postings,” “talked relatively more than they executed problem solving actions,” “monitor results longer,” “produced elaborated plans more frequently” in reference to the tallied codes, frequency of messages and duration of activity. Although the main argument of the paper highlights the authors’ interest in sequential unfolding of regulatory moves, the way the employed quantitative approach isolates and aggregates the actions obscures the temporal connections and sequential mechanisms constituting different forms of regulation moves.

(d) Dillenbourg & Traum (2006) employ a similar methodology to study the relationship between grounding and problem solving in an online environment including a shared whiteboard and a text-chat area. In this study the participants were grouped into dyads and asked to collaboratively work on a murder-mystery task. The authors framed their analysis along the lines of Clark & Brennan’s (1991) theory of grounding (at least applied at the micro level of individual utterances) and theories of socio-cognitive conflict. The study reports two kinds of uses of the dual spaces to facilitate grounding during problem solving: a “napkin” model and a “mockup” model. The authors hypothesized that the whiteboard would be mainly used to disambiguate dialogues in the chat window via basic illustrations (i.e., the napkin model). However, the authors report that the dyads used the whiteboard for organizing factual information as a collection of text boxes, and the chat component was mainly used to disambiguate the information developed on the whiteboard (i.e., the mockup model). The authors attributed this outcome to the nature of the task, which required users to keep track of numerous facts and findings about the murder case, and the difference between the two media in terms of the *persistency* of their contents. Since participants organized key factual information relevant to the problem at hand on the shared whiteboard during their experiments, the authors attributed a shared external memory status to this space and claimed that it facilitated grounding at a broader level by offering a more persistent medium for storing agreed upon facts. The study succeeds in highlighting the important role of medium persistence, even if it does not specify the methods by which students exploited such temporal persistence.

(e) Baker et al. (1999) provide a theoretical account of collaborative learning by bringing together the processes of grounding and appropriation from psycholinguistics and cultural-historical activity theory (CHAT), respectively. In their study they focus on the interactions mediated by the C-Chene software system where dyads are tasked to co-construct energy models that account for storage, transfer and transformation of energy (Baker & Lund, 1997). The models for energy-chains are constructed in a shared workspace that allows the addition of annotated nodes and directed edges. Participants also have access to a chat area that can be customized with sentence openers, which are claimed to promote reflective contributions, reduce typing effort and minimize off-task discussion. The interface is designed to allow only one user to produce a contribution in a given interaction interval. The users need to press a button to switch between dual interaction spaces. Hence the possibility of parallel or overlapping work (e.g., one user drawing on the board as the other is typing a message) is ruled out on the grounds that this would hinder collaboration. The dyads also could not overlap in typing since they need to take turns to use the dialog box where they type their messages. However, it is possible for a user to interrupt his/her partner through a special prompt, which asks whether it is okay to take the turn. If the partner agrees, then the turn is passed to the other user. The study reported that dyads who used the structured interface exhibited more reflective and focused discussion. The authors point to limitations involved with constraining user actions to fixed categories, but they argued that some of the sentence openers they used correspond to generic speech acts that were used for multiple purposes in the course of interaction.

(f) Suthers et al. (2003) investigate the *referential* uses of shared representations in dyadic online discourse mediated by the Belvedere system. This environment has a chat area as well as a shared workspace where dyads can co-construct evidence maps to represent their arguments as a set of categorized textboxes linked to each other (Suthers et al., 2001). The study compares face-to-face and online cases to investigate how dyads use the system as a conversational resource in each case as they work on a shared task that involves developing hypotheses about the spreading of a disease at a remote island. Categories for deictic uses such as finger pointing, cursor-based deixis, verbal deixis and direct manipulation of objects are identified and applied to the session logs. Based on the distributions of these categories for each case, the authors report that dyads in the online case made use of verbal deixis and direct manipulation of shared objects to compensate for the

limitations of the online environment to achieve referential relationships across dual interaction spaces. Moreover, the study reports that such referential links are more likely to be observed between temporally proximal actions. For instance, a chat posting including a deictic term is likely to be read in relation to a node recently added to the shared representation.

Our review of relevant work in the CSCL literature highlights some common threads in terms of methodological approaches and theoretical orientations.¹ First, the studies we have reviewed all focus on the group processes of collaboration, rather than treating it as a mere experimental condition for comparing the individuals in the groups. Second, all studies employ a content-coding approach to categorize actions occurring in multiple interaction spaces. In most cases, representational features like sentence openers or nodes corresponding to specific ontological entities are implemented in the interface to guide/constrain the possibilities for interaction. Such features are also used to aid the categorization of user actions. The categorization schemes are applied to recorded logs and subjected to statistical analysis to elicit interaction patterns.

Despite the accomplishments of these studies, we find that their approaches introduce systematic limitations. Interactional analysis is impossible because coherent excerpts from recorded interactions are excluded from the analysis itself. (Excerpts are frequently used outside of the quantitative analysis, to introduce the features of the system to the reader, to illustrate the categorization schemes employed or to motivate speculative discussion). Moreover, most studies like these involve dyads working on specific problem-solving contexts through highly structured interfaces in controlled lab studies in an effort to manage the complexity of collaboration. The meanings attributed by the researchers to such features of the interface need to be discovered/unpacked by the participants as they put them into use in interaction—and this critical process is necessarily ignored by the methodology. Finally, most of the papers are informed by the psycholinguistic theory of common ground, and are unable to critique it thoroughly.

The Unit of Analysis

For methodological reasons, quantitative approaches generally (a) constrain subject behaviors, (b) filter (code) the data in terms of operationalized variables and (c) aggregate (count) the coded data. These acts of standardization and reduction of the data eliminate the possibility of observing the details and enacted processes of unique, situated, indexical, sequential, social interaction (Stahl, 2006, Ch. 10). An alternative form of interaction analysis is needed to explore the organization of interaction that can take place in CSCL settings.²

In The VMT Project (Stahl, 2009), we also focus on small-group interactions mediated by multimodal interaction spaces. However, our study differs from the work reviewed above by our focus on groups larger than dyads whose members are situated outside a controlled lab environment, and by our use of open-ended math tasks where students are encouraged to come up with their own problems. Moreover, we do not impose any deliberate restrictions on the ways students access the features of our online environment or on what they can say. Our main goal is to investigate how small groups of students construe and make use of the “available features” (Dohn, 2009) of the VMT online environment to discuss mathematics with peers from different schools outside their classroom setting. In other words, we are interested in studying interactional achievements of small groups in complex computer mediations “in the wild.”

Our interest in studying the use of an online environment with multiple interaction spaces in a more naturalistic use scenario raises serious methodological challenges. In an early VMT study where we conducted a content analysis of collaborative problem-solving activities mediated by a standard text-chat tool in a similar scenario of use, we observed that groups larger than dyads exhibit complex interactional patterns that are difficult to categorize based on a theory-informed coding scheme with a fixed/predetermined unit of analysis (Stahl, 2009, Ch 20). In particular, we observed numerous cases where participants post their messages in multiple chat turns, deal with contributions seemingly out of sequence and sustain conversations across multiple threads that made it problematic to segment the data into fixed analytic units for categorization. Moreover, coming to an agreement on a code assignment for a unit that is defined *a priori* (e.g., a chat line) turned out to be heavily dependent upon how the unit can be read in relation to resources available to participants (e.g., the problem description) and to prior units (Stahl, 2009, Ch 22). In other words, the sense of a unit not only depends on the semantic import of its constituent elements, but also on the occasion in which it is embedded (Heritage, 1984). This often makes it possible to apply multiple categories to a given unit and threatens the comparability

¹ We do not intend to minimize the contributions of the particular papers or authors reviewed. On the contrary, we have selected exemplary CSCL studies in order to make a methodological comparison. The quantitative studies may be effective in pursuing their research questions, but their approaches are inadequate for understanding common ground qualitatively. Some of these authors have also adopted case-study approaches more recently; to take only examples from one of the labs, see the studies of deixis, interactional up-take and narrative structure in (Dwyer & Suthers, 2006; Suthers, 2006; Yukawa, 2006).

² It should be clear that we do not reject the use of coding-and-counting for pursuing appropriate research questions—such as testing models of dependent variables. However, they may be systematically unsuited to explore issues of sequential organization and group cognition, which we consider important to CSCL theory and practice.

of cases that are labeled with the same category. More importantly, once the data is reduced to codes and the assignments are aggregated, the sequential relationships among the units are lost. Hence, the coding approach's attempt to enforce a category to each fixed unit without any consideration to how users organize their actions in the environment proved to be too restrictive to adequately capture the interactional complexity of chat (Stahl, 2009, Ch 23). In addition, the inclusion of a shared drawing area in our online environment made the use of a theory-driven coding approach even harder due to increased possibilities for interaction. The open-ended nature of the tasks we use in our study makes it especially challenging to model certain types of actions and to compare them against ideal solutions.

The issue of unit of analysis has theoretical implications. In text chat it is tempting to take a single posting as the unit to be analyzed and coded, because a participant defined this as a unit by posting it as a message and because the chat software displays it as a visual unit. However, this tends to lead the analyst to treat the posting as a message from the posting individual—i.e., as an expression of a thought in the poster's mind, which must then be interpreted in the minds of the post readers. Conversation analysis has argued for the importance of interactions among participants as forming more meaningful units for analysis. These consist of multiple utterances by different speakers; the individual utterances take each other into account. For instance, in a question/answer "adjacency pair" the question elicits an answer and the answer responds to the question. To take a pair of postings such as a question/answer pair as the analytic unit is to treat the interaction within the group as primary. It focuses the analysis at the level of the group rather than the individual. As we just discussed, in online text chat responses are often separated from their referents, so the analysis is more complicated (Fuks, Pimentel & Pereira de Lucena, 2006). In general, we find that the important thing is to trace as many references as possible between chat postings and whiteboard actions in order to analyze the interaction of the group as it unfolds (Stahl, 2009, Ch 26). It is through the co-construction of a rich nexus of such references that the group weaves its shared understanding.

Relatedly, the notion of common ground as an abstract placeholder for registered cumulative facts or pre-established meanings has been critiqued in the CSCL literature for treating meaning as a fixed/denotative entity transcendental to the meaning-making activities of inquirers (Koschmann, 2002). The common ground that supports mutual understanding in group cognition or group problem solving is a matter of semantic references that unfold sequentially in the momentary situation of dialog, not a matter of comparing mental contents (Stahl, 2006, pp. 353-356). Committing to a reference-repair model (Clark & Marshall, 1981) for meaning making falls short of taking into account the dynamic, constitutive nature of meaning-making interactions that foster the process of inquiry (Koschmann et al., 2001).

Given these analytical and theoretical challenges, in studying virtual math teams (Stahl, 2009), we have opted for an alternative to the approaches reviewed above that involve modeling of actions and correct solution paths or treating shared understanding as alignment of pre-existing individual opinions. We focus on the *sequence of actions* in which participants co-construct and make use of *semiotic resources* (Goodwin, 2000) distributed across dual interaction spaces to *do* collaborative problem-solving work. In particular, we focus on the organization of activities that produce graphical drawings on the shared whiteboard and the ways those drawings are used as resources by actors as they collaboratively work on an open-ended math task. Through detailed analysis of a case study (Çakir, Zemel & Stahl, 2009), we investigate how actions performed in one workspace inform the actions performed in the other and how participants coordinate their actions across both interaction spaces. The affordances of the chat and whiteboard spaces are investigated by documenting the methods enacted by participants to address these interactional matters using the affordances of the VMT system (Çakir, 2009). In the next section we will discuss the findings of our case studies in relation to the findings and concerns of related CSCL research that we have covered in our review.

Grounding through Interactional Organization

The coordination of visual and linguistic methods (across the whiteboard and chat workspaces) plays an important role in the establishment of common ground through the co-construction of references between items in the two spaces. Particularly in mathematics—with its geometric/algebraic dual nature—symbolic terms are often grounded in visual presence and associated visual practices, such as counting or collecting multiple units into a single referent (Goodwin, 1994; Healy & Hoyles, 1999; Livingston, 2006; Sfard, 2008; Wittgenstein, 1944/1956). The visually present can be replaced by linguistic references to objects that are no longer in the visual field, but that can be understood based on prior experience supported by some mediating object such as a name—see the discussion of mediated memory and of the power of names in thought by (Vygotsky, 1930/1978; 1934/1986). A more extended analysis of the co-construction of mathematical artifacts by virtual math teams, the complementarity of their visual, semantic and symbolic aspects, their reliance on pre-mathematical practices and processes of reification into concepts are beyond the scope of this paper and require comparison of multiple case studies (see Çakir, 2009). However, for this paper it is important to understand something of how the interactional organization that we have observed in VMT functions to ground the group's understanding of their math objects as shared group achievements.

In CSCL research there has been an explicit interest in studying how affordances of online environments with multiple interaction spaces facilitate *grounding*, and how grounding processes relate to collaborative problem-solving work mediated by such online environments (Baker et al., 1999; Dillenbourg & Traum, 2006). In this section we will discuss the findings of our case studies in relation to the concerns and results reported in prior CSCL research on these issues.

As implied in the OCAF study (Avouris et al., 2003) discussed at the beginning of this paper, investigating grounding and problem-solving processes in online dual-interaction environments like VMT requires close attention to the relationships among actions performed in multiple interaction spaces. Detailed math discussions in the VMT environment present practical challenges for mathematical models that aim to exhaustively capture their relationships. For instance, graphical objects are often layered on top of each other by multiple participants. Despite this combinatoric challenge, a modeling approach can still attempt to capture all possible geometric relationships among these graphical objects in a bottom-up fashion. However, when all chat messages referring to the whiteboard objects are added to the mix, the resulting model may obscure rather than reveal the details of the interactional organization through which group members discuss more complicated mathematical objects by treating a collection of atomic actions as a single entity. Terminology is frequently co-constructed in the chat-and-whiteboard environment specifically in order to be able to refer to complexly defined math objects with a shorthand name, icon or symbol.

The challenges involved with the modeling approach are not limited to finding efficient ways to capture all relationships among actions and identifying meaningful clusters of objects. The figurative uses of the graphical objects present the most daunting challenge for such an undertaking. For instance, a specific drawing may be used as a *gloss* (Garfinkel & Sacks, 1970) to talk about an imagined pattern that grows infinitely and takes the shape illustrated on the whiteboard only at a particular stage. In the absence of a fixed set of ontological elements and constraints on types of actions a user can perform, modeling approaches that aim to capture emergent relationships among semiotic objects distributed across multiple interaction spaces need to adequately deal with the retrospective and prospective uses of language in interaction. Rather than relying upon a generic approach to modeling imposed by the researchers, our interactional approach aims to discover the “model”—or, better, the meaning—that was constructed *by the group*.

In another study discussed earlier, Dillenbourg & Traum (2006) offer the napkin and mockup models in their effort to characterize the relationship between whiteboard and chat spaces. In short, these models seem to describe two use scenarios where one interaction space is subordinated to the other during an entire problem-solving session. The complex relationships between the actions performed across both interaction spaces in our case made it difficult for us to describe the interactions we have observed by committing to only one of these models, as Dillenbourg & Traum did in their study. Instead, we have observed that in the context of an open-ended math task groups may invoke either type of organization, depending upon the contingencies of their ongoing problem-solving work. For instance, during long episodes of drawing actions where a model of some aspect of the shared task is being co-constructed on the whiteboard, the chat area often serves as an auxiliary medium to coordinate the drawing actions, which seems to conform to the mockup model. In contrast, when a strategy to address the shared task is being discussed in chat, the whiteboard may be mainly used to quickly illustrate the textual descriptions with annotations or rough sketches, in accordance with the napkin model. Depending on the circumstances of ongoing interaction participants may switch from one type of organization to another from moment to moment. Therefore, instead of ascribing mockup and napkin models to entire problem-solving sessions, we argue that it would be more fruitful to use these terms as glosses or descriptive categories for types of interactional organizations group members may invoke during specific episodes of their interaction.

Another important observation made by Dillenbourg & Traum is that the whiteboard serves as a kind of shared external memory where group members keep a record of agreed-upon facts. In their study, the dyads were reported to post text notes on the whiteboard to keep track of the information they had discovered about a murder-mystery task. This seems to have led the authors to characterize the whiteboard as a placeholder and/or a shared working memory for the group, where agreed-upon facts or “contributions” in Clark’s sense are persistently stored and spatially organized. As Dillenbourg & Traum observed, the scale of what is shared in the course of collaborative problem solving becomes an important issue when a theory operating at the utterance level like contribution theory (Clark & Marshall, 1981) is used as an analytic resource to study grounding processes that span a longer period of time. Dillenbourg & Traum seem to have used the notion of persistence to extend common ground across time to address this limitation. In particular, they argued that the whiteboard grounds the solution to the problem itself rather than the contributions made by each utterance. In other words, the whiteboard is metaphorically treated as a physical manifestation of the common ground.

In our case studies we have observed that the whiteboard does not simply serve as a kind of shared external memory where the group keeps a record of agreed-upon facts, opinions, hypotheses or conclusions. In our sessions the whiteboard is primarily used to draw and annotate graphical illustrations of geometric shapes, although users occasionally post textboxes on the whiteboard to note formulas they find. While the whiteboard mainly supports visual reasoning and textual discussion or symbolic manipulation occurs chiefly in the chat

stream, actions are carefully, systematically coordinated across the media and integrated within an interactionally organized group-cognitive process. The fact that there are inscriptions posted on the whiteboard does not necessarily mean that all members immediately share the same sense of those graphical objects. The group members did considerable interactional work to achieve a shared sense of those objects that was adequate for the purposes at hand. Hence, the whiteboard objects have a different epistemic status in our case studies than in Dillenbourg & Traum's experiment. Moreover, not all contents of the whiteboard are deemed relevant to the ongoing discussion by the participants. Finally, the sense of previously posted whiteboard objects may be modified or become evident as a result of current actions (Suchman, 1990). In other words, group members can not only reuse or reproduce drawings, but they can also make subsequent sense of those drawings or discard the ones that are not deemed relevant anymore. Therefore, the technologically extended notion of common ground as a placeholder for a worked-out solution suffers from the same issues stated in Koschmann & LeBaron's (2003) critique of Clark's theory. As an abstract construct transcendental to the meaning-making practices of participants, the notion of common ground obscures rather than explains the ways the whiteboard is used as a resource for collaborative problem solving.

From Common Ground to Indexical Ground

Instead of using an extended version of common ground as an analytical resource, we frame our analysis using the notion of "indexical ground of deictic reference"—a term we appropriated from linguistic anthropology (Hanks, 1992). In face-to-face interaction, human action is built through the sequential organization of not only talk but also coordinated use of the features of the local scene that are made relevant via bodily orientations, gesture, eye gaze, etc. In other words, "human action is built through simultaneous deployment of a range of quite different kinds of semiotic resources" (Goodwin, 2000, p. 1489). Indexical terms and referential deixis play a fundamental role in the way these semiotic resources are interwoven in interaction into a coherent whole.

Indexical terms are generally defined as expressions whose interpretation requires identification of some element of the context in which it was uttered, such as who made the utterance, to whom it was addressed, when and where the utterance was made (Levinson, 1983). Since the sense of indexical terms depends on the context in which they are uttered, indexicality is necessarily a relational phenomenon. Indexical references facilitate the mutually constitutive relationship between language and context (Hanks, 1996). The basic communicative function of indexical-referentials is "to individuate or single out objects of reference or address in terms of their relation to the current interactive context in which the utterance occurs" (Hanks, 1992, p. 47).

The specific sense of referential terms such as *this*, *that*, *now*, *here* is defined locally by interlocutors against a shared indexical ground. The intelligibility of such terms is based on a *figure/ground relationship* between the new action involving the indexical reference (i.e. the figure) and the sequentially unfolding context constituted by the actions that has been witnessed recently and the objects persistently available in the shared scene. Conversely, the linguistic labels assigned to highlighted features of the local scene shape the indexical ground. Hence, the indexical ground is not an abstract placeholder for a fixed set of registered contributions. Rather, it signifies an emergently coherent field of action that encodes an interactionally achieved set of background understandings, orientations and perspectives that make references intelligible to interlocutors (Zemel et al., 2008).

Despite the limitations of online environments for supporting multimodality of embodied interaction, participants make substantial use of their everyday interactional competencies as they appropriate the features of such environments to engage with other users. For instance, Suthers *et al.*'s (2003) study reports that deictic uses of representational proxies play an important role in the interactional organization of online problem-solving sessions mediated by the Belvedere system. The authors report that participants in the online case devised mechanisms that compensate for the lack of gestural deixis with alternative means, such as using verbal deixis to refer to the most recently added text nodes and visual manipulation of nodes to direct their partner's attention to a particular node in the shared argument map.

In contrast to the Belvedere system, VMT offers participants additional resources such as an explicit referencing mechanism, a more generic workspace that allows producing and annotating drawings, and an awareness feature that produces a sense of sequentiality by embedding indicators for drawing actions in the sequence of chat postings. Our case study (Çakir et al., 2009) shows that despite the online situation's lack of the familiar resources of embodied interaction, team members can still achieve a sense of shared access to the meaningful objects displayed in the dual interaction spaces of the VMT environment. Our analysis indicates that coherence among multiple modalities of an online environment like VMT is achieved through members' methodical uses of the features of the system to coordinate their actions in the interface. In particular, we observed that the witnessable details of the orderly construction of shared inscriptions (e.g., the way objects are spatially arranged in relation to each other through sequences of actions) and the deictic references that link chat messages to features of those inscriptions and to prior chat content are instrumental in the achievement of indexical symmetry (intersubjectivity) with respect to the semiotic objects relevant to the task at hand.

Through coordinated use of indexical-referential terms and highlighting actions, team members help each other to literally “see” the objects implicated in the shared visual field (Goodwin, 1994) and to encode them with locally specified terminology for subsequent use. Moreover, the integration of both modalities in this manner also facilitates joint problem solving by allowing group members to invoke and operate with multiple realizations—graphical, narrative and symbolic—of their mathematical task. Such coordinated work across modalities can be a powerful problem-solving resource since it allows participants to invoke various mathematical practices relevant to the task at hand and to make use of them in mutually elaborating ways.

To sum up, the focus of our ethnomethodological inquiry is directed towards documenting how virtual teams achieve a sense of reciprocity and coherence among their actions in an online CSCL environment with multiple interaction spaces. Our approach (Stahl, 2009, Ch 28) includes a close investigation of the moment-to-moment details of the practices through which participants organize their chat utterances and whiteboard actions as a coherent whole in interaction—a process that is generally lost in statistical analyses of multiple cases, where categorization and aggregation systematically miss the rich and vital relationships of indexicality and sequentiality. We have observed that referential practices enacted by the users are essential in the coordinated use of multimodalities afforded by such environments. The referential uses of available features are instrumental not only in allocating other members’ attention to specific parts of the interface where relevant actions are being performed, but also in the achievement of reciprocity (intersubjectivity, common ground, shared understanding, group cognition) among actions in the multiple interaction spaces, and hence a sense of sequential organization across the spaces. Among the things that count in interaction are the indexical ground that is built up through interactional references that cannot be counted without destroying their sequential relationships.

References

- Avouris, N., Dimitracopoulou, A., & Komis, V. (2003). On analysis of collaborative problem solving: An object-oriented approach. *Computers in Human Behavior, 19*, 147-167.
- Baker, M., Hansen, T., Joiner, R., & Traum, D. (1999). The role of grounding in collaborative learning tasks. In P. Dillenbourg (Ed.), *Collaborative learning: Cognitive and computational approaches* (pp. 31-63). Oxford, UK: Pergamon.
- Baker, M., & Lund, K. (1997). Promoting reflective interactions in a CSCL environment. *Journal of Computer Assisted Learning, 13*, 175-193.
- Çakir, M. P. (2009). *How online small groups co-construct mathematical artifacts to do collaborative problem solving*. Unpublished Dissertation, Ph.D., College of Information Science and Technology, Drexel University, Philadelphia, PA, USA.
- Çakir, M. P., Zemel, A., & Stahl, G. (2009). The joint organization of interaction within a multimodal CSCL medium. *International Journal of Computer-Supported Collaborative Learning (ijCSCL), 4*(2).
- Clark, H., & Brennan, S. (1991). Grounding in communication. In L. Resnick, J. Levine & S. Teasley (Eds.), *Perspectives on socially-shared cognition* (pp. 127-149). Washington, DC: APA.
- Clark, H. H., & Marshall, C. (1981). Definite reference and mutual knowledge. In A. K. Joshi, B. Weber & I. A. Sag (Eds.), *Elements of discourse understanding* (pp. 10-63). New York, NY: Cambridge University Press.
- Dillenbourg, P. (2005). Dual-interaction spaces. In T. Koschmann, D. D. Suthers & T.-W. Chan (Eds.), *Computer-supported collaborative learning 2005: The next ten years! (Proceedings of CSCL 2005)*. Taipei, Taiwan: Mahwah, NJ: Lawrence Erlbaum Associates.
- Dillenbourg, P., & Jermann, P. (2006). Designing integrative scripts. In F. Fischer, H. Mandl, J. Haake & I. Kollar (Eds.), *Scripting computer-supported collaborative learning: Cognitive, computational and educational perspectives* (pp. 275-301). Dordrecht, Netherlands: Kluwer-Springer Verlag.
- Dillenbourg, P., & Traum, D. (2006). Sharing solutions: Persistence and grounding in multimodal collaborative problem solving. *Journal of the Learning Sciences, 15*(1), 121-151.
- Dohn, N. B. (2009). Affordances revisited: Articulating a Merleau-Pontian view. *International Journal of Computer-Supported Collaborative Learning (ijCSCL), 4*(2).
- Dwyer, N., & Suthers, D. (2006). Consistent practices in artifact-mediated collaboration. *International Journal of Computer-Supported Collaborative Learning, 1*(4), 481-511. Retrieved from <http://dx.doi.org/10.1007/s11412-006-9001-1>.
- Fuks, H., Pimentel, M., & Pereira de Lucena, C. (2006). R-U-Typing-2-Me? Evolving a chat tool to increase understanding in learning activities. *International Journal of Computer-Supported Collaborative Learning, 1*(1), 117-142. Retrieved from <http://dx.doi.org/10.1007/s11412-006-6845-3>.
- Garfinkel, H., & Sacks, H. (1970). On formal structures of practical actions. In J. Mckinney & E. Tiryakian (Eds.), *Theoretical sociology: Perspectives and developments* (pp. 337-366). New York, NY: Appleton-Century-Crofts.
- Goodwin, C. (1994). Professional vision. *American Anthropologist, 96*(3), 606-633.
- Goodwin, C. (2000). Action and embodiment within situated human interaction. *Journal of Pragmatics, 32*, 1489-1522.
- Hanks, W. (1992). The indexical ground of deictic reference. In C. Goodwin & A. Duranti (Eds.), *Rethinking context: Language as an interactive phenomenon*. Cambridge, UK: Cambridge University Press.
- Hanks, W. (1996). *Language and communicative practices*. Boulder, CO: Westview.
- Healy, L., & Hoyles, C. (1999). Visual and symbolic reasoning in mathematics: Making connections with computers. *Mathematical Thinking and Learning, 1*(1), 59-84.
- Heritage, J. (1984). *Garfinkel and ethnomethodology*. Cambridge, UK: Polity Press.

- Jermann, P. (2002). *Task and interaction regulation in controlling a traffic simulation*. Paper presented at the Computer support for collaborative learning: Foundations for a CSCL community. Proceedings of CSCL 2002, Boulder, CO. Proceedings pp. 601-602.
- Jermann, P., & Dillenbourg, P. (2005). Planning congruence in dual spaces. In T. Koschmann, D. D. Suthers & T.-W. Chan (Eds.), *Computer-supported collaborative learning 2005: The next ten years! (Proceedings of CSCL 2005)*. Taipei, Taiwan: Mahwah, NJ: Lawrence Erlbaum Associates.
- Komis, V., Avouris, N., & Fidas, C. (2002). Computer-supported collaborative concept mapping: Study of synchronous peer interaction. *Education and Information Technologies*, 7(2), 169-188.
- Koschmann, T. (2002). Dewey's contribution to the foundations of CSCL research. In G. Stahl (Ed.), *Computer support for collaborative learning: Foundations for a CSCL community: Proceedings of CSCL 2002* (pp. 17-22). Boulder, CO: Lawrence Erlbaum Associates.
- Koschmann, T., & LeBaron, C. (2003). *Reconsidering common ground: Examining clark's contribution theory in the operating room*. Paper presented at the European Computer-Supported Cooperative Work (ECSCW '03), Helsinki, Finland. Proceedings pp. 81-98.
- Koschmann, T., LeBaron, C., Goodwin, C., & Feltovich, P. J. (2001). 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.
- Levinson, S. (1983). *Pragmatics*. Cambridge, UK: Cambridge University Press.
- Livingston, E. (2006). Ethnomethodological studies of mediated interaction and mundane expertise. *The Sociological Review*, 54(3).
- Mühlpfordt, M., & Wessner, M. (2005). Explicit referencing in chat supports collaborative learning. In T. Koschmann, D. D. Suthers & T.-W. Chan (Eds.), *Computer-supported collaborative learning 2005: The next ten years! (Proceedings of CSCL 2005)* (pp. 460-469). Taipei, Taiwan: Mahwah, NJ: Lawrence Erlbaum Associates.
- Neuendorf, K. A. (2002). *The content analysis guidebook*. Thousand Oaks, CA: Sage.
- Roschelle, J., & Teasley, S. (1995). The construction of shared knowledge in collaborative problem solving. In C. O'Malley (Ed.), *Computer-supported collaborative learning* (pp. 69-197). Berlin, Germany: Springer Verlag.
- Scardamalia, M., & Bereiter, C. (1996). Computer support for knowledge-building communities. In T. Koschmann (Ed.), *CSCL: Theory and practice of an emerging paradigm* (pp. 249-268). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Sfard, A. (2008). *Thinking as communicating: Human development, the growth of discourses and mathematizing*. Cambridge, UK: Cambridge University Press.
- Soller, A. (2004). Understanding knowledge sharing breakdowns: A meeting of quantitative and qualitative minds. *Journal of Computer Assisted Learning*, 20, 212-223.
- Soller, A., & Lesgold, A. (2003). *A computational approach to analyzing online knowledge sharing interaction*. Paper presented at the 11th International Conference on Artificial Intelligence in Education, AI-ED 2003, Sydney, Australia. Proceedings pp. 253-260: Amsterdam: IOS Press.
- Stahl, G. (2006). *Group cognition: Computer support for building collaborative knowledge*. Cambridge, MA: MIT Press. Retrieved from <http://GerryStahl.net/mit/>.
- Stahl, G. (Ed.). (2009). *Studying virtual math teams*. New York, NY: Springer. Computer-supported collaborative learning book series, vol 11 Retrieved from <http://GerryStahl.net/vmt/book>.
- Suchman, L. A. (1990). Representing practice in cognitive science. In M. Lynch, Woolgar, S. (Ed.), *Representation in scientific practice*. Cambridge, MA: MIT Press.
- Suthers, D. (2006). Technology affordances for intersubjective meaning making: A research agenda for CSCL. *International Journal of Computer-Supported Collaborative Learning*, 1(3), 315-337. Retrieved from <http://dx.doi.org/10.1007/s11412-006-9660-y>.
- Suthers, D., Connelly, J., Lesgold, A., Paolucci, M., Toth, E., Toth, J., et al. (2001). Representational and advisory guidance for students learning scientific inquiry. In K. D. Forbus & P. J. Feltovich (Eds.), *Smart machines in education: The coming revolution in educational technology* (pp. 7-35). Menlo Park: AAAI Press.
- Suthers, D., Girardeau, L., & Hundhausen, C. (2003). Deictic roles of external representations in face-to-face and online collaboration. In B. Wasson, S. Ludvigsen & U. Hoppe (Eds.), *Designing for change in networked learning environments, Proceedings of the international conference on computer support for collaborative learning 2003* (pp. 173-182). Dordrecht: Kluwer Academic Publishers.
- Vygotsky, L. (1930/1978). *Mind in society*. Cambridge, MA: Harvard University Press.
- Vygotsky, L. (1934/1986). *Thought and language*. Cambridge, MA: MIT Press.
- Wittgenstein, L. (1944/1956). *Remarks on the foundations of mathematics*. Cambridge, MA: MIT Press.
- Yukawa, J. (2006). Co-reflection in online learning: Collaborative critical thinking as narrative. *International Journal of Computer-Supported Collaborative Learning*, 1(2), 203-228. Retrieved from <http://dx.doi.org/10.1007/s11412-006-8994-9>.
- Zemel, A., Koschmann, T., LeBaron, C., & Feltovich, F. (2008). "What are we missing?" Usability's indexical ground. *Computer Supported Cooperative Work*, 17, 63-85.