Context Aware CSCL: Moving Toward Contextualized Analysis

Sean P. Goggins, Drexel University, outdoors@acm.org
James Laffey, University of Missouri – Columbia, laffeyj@missouri.edu

Abstract: Groups change over time. CSCL groups, unlike face to face groups, leave evidence of their interactions behind in the form of system logs. To this point, much analysis of system logs has been opportunistic; researchers and tool developers make visible what is available. A number of specific empirical studies have pointed to the promise of using log data for process modeling, stochastic modeling, dependency graph construction and other strategies for unpacking and explaining the interactive, and dynamic relationships that unfold in CSCL systems. We have been studying a system supporting context awareness and learning in completely online CSCL groups for four years using a new type of logging infrastructure intended to increase participant awareness of their social context. Analysis of this and related data has resulted in a number of published studies pointing to the potential for understanding how groups emerge, change and disband in completely online CSCL courses. These publications, and our research program more broadly have led us to an important socio-technical insight, and a possible theoretical contribution. The socio-technical insight is that logs can be designed, and should be designed to capture the data that helps to support both member awareness in a socio-technical environment and the needs of researchers to better understand collaborative learning interactions. We use CANS logs as a foundation for analysis, and explain the restructuring of those logs to reflect the online social experience of users more accurately. We then apply this socio-technical insight to advocate for exploring CSCL group interaction from a more socially focused theoretical perspective.

Introduction

In this concept paper we assert that theories of interaction in CSCL must incorporate a more explicit approach to the design of logging systems. Suthers (2010) describes contingency graphs as analytical mechanisms for the study of uptake in CSCL environments in a manner that does not place time at the center of analysis, but instead focuses on conditions that precede and follow important learning acts. Reimann (2009) focuses on the central importance of time, highlighting a gap in our understanding of how time is understood in long running, asynchronous interaction. Stahl (2006; Stahl, 2009b; Stahl, 2009b) focuses on the small group unit of analysis as he calls for a study of the science of group interaction (Stahl, 2009a). Our contribution exists at the intersection of these interests and efforts. We integrate Stahl's focus on the small group unit of analysis, Reimann's attention to time, Suther's analytic view of data and four years of experience using rich, context aware logs to study asynchronous, completely online CSCL environments (cited throughout). With the rest of this paper we review the literature in CSCL focused on advancing the analytical tools available in the field, propose a framework for more closely connecting interactions to the lived experience of users, and discuss the implications for CSCL theories.

Literature Review

Suthers (2006) called for recognition that ongoing work in CSCL requires integration of design based, phenomenological, and experimental methods to build a complete picture of the intersubjectivity of computer supported collaborative learning. Each methodological tradition is, by itself, too narrow to support the ongoing examination of the socio-technical CSCL experience that we now recognize to be munificent in its variation, even in the same population using the same set of tools over time. Specifically, Suthers (2006) points out that experimental approaches embody an artificiality that constrains our view of how learning events actually happen in the world, design based approaches take the form of iterations and lead to an emergent technomethodology (Dourish & Button, 1996) that practically celebrates munificent variation in uptake and use of CSCL ICT's, and more purely phenomenological studies of CSCL provide descriptions of what occurred. These approaches often fall short in their attempts to provide useful guidance for the development of interventions in the future.

Reimann (2009) takes another perspective on the range of methods utilized in CSCL by presenting a contrast between coding and counting CSCL events, which he calls variable focused analysis, and sequential analysis of events in CSCL settings. This work raises questions about the applicability of analysis of variables in the complex, real world settings that constitute most CSCL environments and the intersubjective learning processes they support. The observed munificent variations of CSCL experience are, to Reimann, an irreconcilable set of multivariate confounds.

As a solution this to core challenge in CSCL research, Reimann (2009) posits that analysis of events and event streams will provide a more authentic view of the intersubjective nature of learning in CSCL environments. While the conceptual insight about the importance of event logs is one we agree with, the conclusion that these events are likely to be digested into semantically meaningful process models and process model instances presumes a linearity and consistency of interaction in CSCL environments for which there is no empirical evidence. Reimann, Frerejean & Thompson (2009) test the idea of applying a process model to event data, concluding that the decision processes in fact take a different path each time. Goggins, Laffey & Tsai (2007), Goggins, Laffey & Galyen (2010a) & Goggins et al (2010b) go further in demonstrating the munificent variations of interaction revealed by event logs. These studies show that even in the most controlled CSCL environments, no two groups follow the same processes. Specifically, those CSCL participants who experience the same CSCL curriculum in the same CSCL, socio-technical environment with the same instructor produce interaction logs with highly variable activity levels, activity sequences and following group-specific processes. This holds for groups who are able to see the activity of other groups and would therefore be at least partially susceptible to social comparison influences (Festinger, 1954) and those where groups are not able to see each others activity (Goggins et al., 2011, International Journal of Computer Supported Cooperative Work, Under Reveiew). Put simply, the notion of nascent, identifiable process models emerging across instances of a CSCL environment is not supported by any data we are aware of.

CSCL in the wild is not experienced as set linear patterns that fit into process models, but instead as a set of time bounded social groupings within the socio-technical context (Goggins et al., 2010a). The gap between findings about the social nature of CSCL, compared to findings asserting a more process oriented nature in CSCL experience relies on two key premises in the process oriented work. First, the importance of indexicality in CSCL communication logs. Indexicality is a construct that claims the sequence of events is significant for understanding each event. What happens before and after an action matters. A large corpus of literature in CSCL supports this claim (Olson, Herbsleb, & Rueter, 1994; Bruckman, 2002; Cakir, 2007; Suthers, Dwyer, Medina, & Vatrapu, 2007; Laffey, Amelung, & Goggins, 2009; Reimann, 2009; Reimann et al., 2009; Stahl, 2009b; Stahl, Ou, Cakir, Weimar, & Goggins, 2010), and this premise is not contradicted by the social theories of learning within which our work is framed. The second premise underlying the process modeling approach is that consistent processes can be observed across multiple CSCL instances. This premise is antithetical to social theories of learning, which assert that our knowledge is measured by our ability to contribute to socially constructed practices in our communities (Wenger, 1998). Social construction is an emergent, natural process, while process models rely on a conceptualization of a priori order being replicated to some degree across multiple socio-technical instances of an activity.

Process modeling approaches to CSCL event data analysis seek to place the munificent variation of activity in socio-technical CSCL systems into a box that mutes the variation, and artificially constrains analysis of event log data to linear sequences of activity. Analysis of events that take the social nature of CSCL experience into account will need to consider non-linear and non-scripted events and social structures as well. We assert that process models are useful in the narrow case of scripted CSCL interactions, but swiftly lose utility when applied to emergent group interactions in a socio-technical system. Viewed from the point of view of non-scripted activities, CSCL process models seem likely to either be too abstract to provide useful guidance, or too detailed to expect the development of common patterns. Reimann (2009) acknowledges that event streams from a socio-technical environment do not necessarily result in clear semantics for the activities logged and that the indexicality of a single set of interactions is not likely to be repeated across multiple instances of CSCL interaction.

An alternative to process modeling is Suthers et al's (2007) concept of an eclectic model for examining the interactivity of participants in a CSCL environment. Unlike process models, the eclectic model works to incorporate multiple perspectives from different data. The events themselves are not any more richly constructed than Reimann's (2009), but they are integrated with other data to tell the story of CSCL experience from multiple methodological perspectives. Suthers et al (Suthers et al., 2007) explicate the construct of a dependency graph, which they use as an analytical boundary object for integrating event log data with data from other data types and research methods commonly used in CSCL research. Like Reimann, Suthers et al arrive at a method of examining behavior in logs that relies on establishing a more defined, non-dynamic picture of interaction than what is born out by much experience in the socio-technical systems that constitute CSCL in the wild. The important contribution of the dependency graph in CSCL research is that it facilitates consistent integration of data from the diverse set of research traditions used to examine CSCL.

Eclectic modeling and dependency graphs used as boundary objects raise the question of whether or not existing systems for logging CSCL activity provide sufficient contextual data to support an automated approximation of a dependency graph. Is it possible more context data could be captured in logs than is commonly the case in CSCL today? If we did this, could the automated generation of dependency graphs that are more easily integrated with other forms of CSCL data be realized? The interactive logs we analyze provide the bidirectional view of interaction through technology in a learning environment that is not available for the analysis methods proposed by Reimann (Reimann et al., 2009) and Suthers (Suthers et al., 2007). The logs produced by the CANS system (our unique logging system, described under study context and methods), in fact, provide a wider swath of interaction data for analysis, and as a consequence greater potential for automated analysis of interactivity in computer mediated learning settings. For example, our preliminary analysis of interaction logs over three years of CANS data shows that these sorts of passive, invisible interactions between members are more common than post behaviors by a ratio of 15:1 (6.6% of activity in ~670,000 events is active, posting or creating activity). To understand the uptake of ideas in these environments, CANS logs provide a wide foundation of interactivity records that we will use for the development of insight from events, event vocabularies and event grammars (event vocabularies and grammars are discussed in more depth later in this section).

Our work seeks to elaborate on Suthers et al's (2007) notion of the dependency graph by incorporating a more complete logging infrastructure, and analysis that recognizes the evolving social structures and patterns that can be made visible through these logs. Most significantly, the logs we use capture both passive (reading) and active (posting) behavior of participants. Both Reimann & Suthers et al's approaches rely on the analysis of interactions and interaction logs that only contain a record of the proactive posting behavior of participants. This proactive posting behavior, in response to others and sometimes starting from scratch (as in a new forum) describes the observable, creative acts of participants. Prior research in online awareness (Carroll, Neale, Isenhour, Rosson, & McCrickard, 2003; Carroll, Rosson, Convertino, & Ganoe, 2006; Amelung, 2007; Laffey et al., 2009) across multiple contexts suggests that knowledge of the social presence (Erickson & Kellogg, 2000) of others influences interaction. In CSCL environments, knowledge of who is reading the contributions of which other participants permits the researcher to observe the full intersubjective nature of interactions, and tease out the vocabularies and grammars of interaction which correspond with different levels of performance.

Vocabularies and grammars of interaction can be constructed from the context aware logs we use. A vocabulary is a list of event types captured in a socio-technical environment. An event grammar is an arrangement of event types in linear or possibly non-linear (non-sequential, non-indexical) groupings that provide a heuristic indication of the trajectory of activities in the socio-technical environment. This context aware record of interaction in the online environment provides more information than is otherwise available to the researcher. For example, as we noted, dependency graphs are boundary objects for the integration of research methods and data across the range of CSCL methods, including iterative design, phenomenological work and experiments. To advance CSCL research that incorporates these views, we need a logging system that captures context for a defined vocabulary of events. Those events may then be used, in conjunction with other data, to construct CSCL event grammars.

Conceptually, grammars are micro-patterns that can be arranged like lego blocks in socio-technically embodied interactions. Unlike process models, these patterns can be centered on individual connections (social networks) or topical clusters and social clusters. The context metadata, event vocabularies and bi-directional nature of the event logs generated by CANS make this possible. Bi-directional, context aware log analysis is an instrument that enriches and accelerates the construction of dependency graphs. Such logs act as an activity index to reference across all other analysis of data.

To begin identifying event semantics using these grammatical building blocks, we start with Suthers et al's (2007) conceptualization of uptake as an interpretive act. We then proceed to describe how we analyze the bi-directional sequences and socio-technical contexts through which objects are brought forth and provided with further meaning through interaction. For Suthers et al's data, this interpretive act focuses on the artifacts that are visible at the time of analysis, and the active behaviors of participants acting against those objects. Information about the more passive, reading behaviors of participants in these systems is not taken into account in Suthers' analysis of uptake.

Our work views this less active behavior as a significant but typically invisible indicator of uptake in CSCL environments. As an example, imagine an individual in an online graduate student course who is contributing to a discussion in her small group. The participation in this course occurs in course wide discussion boards and through discussion boards restricted to each small group. Our subject is participating in a design activity with her small group while at the same time contributing to a question posed by the instructor of the course, asking students to list a design researcher whose work is exemplary of the type of work she would like to do. If we know that the student is reading the posts of other students in the larger context consistently before she makes contributions to her small group, we have evidence that the meaning making (uptake) is influenced not only by her direct response to her group, but also by her sequential review of other material in the course. If we extend our view of the uptake of information to include other interactions of the member with information resources or colleagues outside the course, it quickly becomes apparent that our view of uptake is materially influenced by the width of data we capture.

Our work adapting Suthers et al's (2007) notion of uptake emerges as a two-step interpretive dance. The first step is analyzing sets of log data that permit a researcher to witness emergent social structures, combined with the full indexicality of socio-technical interactions – active and passive. This will potentially change both our view of the nature of mixed methods research in CSCL, but also expose the potential richness of member event grammars that may be constructed from these logs. As we noted earlier, passive actions outnumber active actions in the systems we have studied by a ratio of 15:1. Our logs provide 15x more data for constructing event grammars. As a second step outside the scope of this paper, but important to our long term research agenda, capturing log data from multiple socio-technical contexts to reflect the actions of a group of CSCL users will widen our view of indexicality and possibly introduce a new set of CSCL patterns.

Prior work with conversation analysis and structural analysis in the social sciences – our core theoretical and methodological frames for this work – all point to the small group as the most salient unit of analysis for our work. Unit of analysis is an important consideration in any CSCL research, and it is especially important when the research questions are focused on the analysis of interaction. Individuals, small groups and larger classrooms and communities are the common units of analysis in CSCL. We have log data that might be viewed from any of these perspectives. Our data and our analysis, however, are targeted at the small group unit of analysis. This unit of analysis permits us to examine the interactivity among members of small teams with each other, as well as their behavior in a larger socio-technical context. In this way, the small group unit of analysis truly becomes, as Stahl (2006) states, "Where the action is" in CSCL.

As the reader will notice through our methods and analysis, we are looking for patterns in log data. We proceed with the rest of this paper by first unpacking what an event is in an asynchronous CSCL environment. We accomplish this through description of prior empirical and theoretical work examining deictic structure in synchronous CSCL interactions. Then, we show how CANS events are captured and analyzed to provide a view of the emergent social and interaction structures, combined with the full indexicality of socio-technical interactions in an asynchronous CSCL environment. Second, we describe the methods and context of this study. Third, we present our findings. Fourth, we discuss the implications of our findings for future versions of bidirectional CSCL logging with CANS and mixed methods CSCL analysis in general. Finally, we describe the next steps we see as critical in CSCL research programs that leverage log data.

Which Log Structures are Analytically Useful?

There are wide ranges of tradition for studying interaction in socio-technical and physically situated environments that inform our research. Conversation analysis from the ethnomethodological tradition and social network analysis from structural sociology contribute salient theoretical and methodological perspectives that inform our work to advance log analysis in CSCL. The context data and completeness of the interaction record that is revealed by our log data is also foundational to our ability to proceed with a study that explores the utility of log and sequence structures for CSCL research.

Research methods like conversation analysis have examined the microstructure and indexicality of conversation to explicate meaning. Our goal here is to discover new ways of analyzing electronic traces by identifying semantically meaningful grammars from event trace data in CSCL environments. The events we record and grammars we identify are derived from socio-technical interactions recorded in CANS event logs from an online course management system. While conversation analysis has a long tradition of examining micro-sequences of conversation to explicate meaning and understand the perspective of conversational participants, our method is new. In conversation analysis, the sequential order of utterances and non-verbal communication are viewed together with a particular expression, word or phrase to discern meaning. The presence or absence of eye contact, patterns of turn taking and changes in body posture are all meta-factors that conversation analysis researchers take into consideration when discerning meaning from speech acts and interactions in the physical world. In CSCL, the Virtual Math Teams research led by Gerry Stahl (Stahl, 2006; Cakir, 2007; Stahl, 2009b; Stahl et al., 2010) demonstrate the applicability of methods derived from conversation analysis and ethnomethodology to the analysis of socio-technical conversations around objects in a synchronous CSCL space. As in conversation analysis, Stahl's approach to describing the emergence of group cognition during synchronous collaborative learning relies on micro-analysis of, in his case, socio-technical interactions in a tool specially designed to support math discourse.

Conversation analysis requires a significant investment of researcher time, of course. The successful adaptation of conversation analysis and the construction of ethnomethodologically informed research methods centered on a socio-technical system like VMT naturally begs the question: Is there a way to automate the analysis of electronic trace data in order to arrive at some reasonable approximation of the hard won findings that emerge from small groups of researchers discussing small group conversations? After all, unlike real life, we have electronic interaction traces readily available.

Readers should note that we think that obtaining the richness of understanding that emerges from ethnomethodolically informed methods, like those used on the VMT project through automated means, is an unrealistic goal. Without the benefit of close, interpretive examination of discourse there is certain to be loss in meaning, which is a natural outcome of choosing a vocabulary of events and event grammars, which are required steps in the production and analysis of electronic trace data. Semantically meaningful events, as

Reimann (2009) observes, are constructed from the logged actions of users. Our goal is an improvement in automated analysis through a rich type of logging and the connection of that analysis to actual member performance. The success of ethnomethodologically informed methods in CSCL research inspires our work. We view the types of event vocabularies and grammars developed through our work as an accelerator for the integration of multiple types of CSCL data, elaborating Suthers notion of a dependency graph.

To begin to accomplish an approximation of the understanding derived through conversation analysis from automated analysis of logs, we need to identify meaningful structures and sequences (grammars) from the available logged events (vocabularies). The construction of semantically meaningful events begins with a vocabulary of logged event types, and a search for social structural patterns (who relates to who), the trajectory of evolution in social structure and a complementary analysis of interaction sequences across the socio-technical context, which includes the layers of the system context. Events, however they are defined, are initially captured at some imperfect, but consistent granularity that works to mute subtle gestures, references and turn-taking activities that constitute the actual, qualitative experience of conversation in the electronically mediated or real world.

Event grammars built up from event vocabularies do not hold enough information to permit us to approximate the results of conversation analysis research methods through automated log analysis. We must also connect performance within groups to the structural patterns and sequences of interaction made visible through our bidirectional usage logs in order to connect events to our understanding of uptake in CSCL.

Social Network Analysis of Sequences of Interaction

Network theory provides a particularly helpful frame for understanding the relationships between individuals and groups as they develop. This type of analysis holds great potential for the study of CSCL groups, particularly where online activity logs are available. Unlike network analysis of physically situated groups, which has been criticized for relying on data with low reliability, online groups for whom use logs are kept may be analyzed using a computer record of their activities which can be reliably recorded and holds specific meanings. Network theory rests on five core positions (Katz, Lazer, Arrow, & Contractor, 2004):

1. Behavior is best predicted by examining the relations people find themselves embedded in,
2. The focus of analysis should be the relationships between units in a group
3. Analytical methods must assume that there is interdependence among individuals in a network, so normal assumptions of independence are not valid
4. Understanding a social system requires analysis of the structure of an entire network, not simply the ties between two members (dyadic ties)
5. Group relations are fuzzy, which means that where a specific group's boundaries are, and whether a particular individual is a member of a particular group isn't clear-cut. Group assignment requires some judgment on the part of the researcher. Fortunately, there is ample prior work to use as a guide for such judgments.

Social network analysis (SNA) is the method researchers use to guide empirical works framed by network theory. SNA has numerous specific forms, each informed to different degrees by graph theory, probability theory and algebraic modeling. The essential concepts represented by SNA are the actor – who can be an individual, group or event, the relational tie (between actors), dyads (two people), triads (three people), groups, subgroups and networks. Next, we will briefly review the core concepts used in SNA, describe the small number of studies which have used SNA to describe online group development and review more computationally focused work that addresses the analysis of communication patterns and social networks from a myriad of online network data types.

Socio-Centric density in a valued network describes the total of all actual connections divided by the number of possible ties, resulting in an average value. If values (number of connections between nodes) are dichotomized, important data is removed (connection strength between nodes), but the resulting 0-1 numbers provide a measure of network completeness that is easier to compare with other networks. Network centralization measures indicate how tightly the graph (social network) is organized around its most central point. High in degree centralization indicates that the group is focused inward on a few core members. High out degree centralization means that a few core members are producing most of the connections to others. Betweenness is a measure of the importance of a node to making connections between other nodes. For example, if there are two clusters of individuals in a social setting, a person with high betweenness would be a member of both clusters. This would indicate that they are a "connection point" for ideas between two clusters within the larger group. Core-periphery analysis extends the established centrality SNA measures of degree, closeness, betweenness and eigenvector-based measures by leveraging the concepts of group centrality and two-mode centrality (Wasserman & Faust, 1994). Core/periphery measures posit that there is some group in the core, and some other group in the periphery of a network with one center. Core nodes are distinguished by creating a complete graph with each other.

SNA contributes an important, orthogonal perspective to our analysis of context enriched, bi-directional event data. Earlier, we described how event vocabularies and grammars derived from event logs could be used to make the indexicality of socio-technically mediated conversational moves visible to researchers more quickly. SNA adds the perspective of the networked relationships of individuals in small groups. We can focus this lens on short interaction periods of days or weeks, or longer interaction periods of months to discern social connectedness among participants systematically, or to understand structure during different, specific types of interaction sequences. SNA complements the construction of event grammars with insight into who is connected to whom in a CSCL context.

Measuring Performance in CSCL: Content Analysis

Social connections that emerge from SNA and micro-level event grammars that emerge from our logs help us to see the structure and sequence of activity. The relation of those structures and sequences to performance and outcomes in CSCL is a vital step for making claims about the significance of these grammars and structures for CSCL research. One core claim of Stahl's (Stahl, 2006) work is that group cognition is a construct that emerges from the observable interactions of group members. Through various analytical strategies, ideas that cannot be solely attributed to a single individual or artifact become visible. Conversation analysis is salient for synchronous activity like that which Stahl examines. The corollary for asynchronous CSCL is content analysis, which we use to classify the degree of knowledge construction in the asynchronous discussion boards in online CSCL.

Gunawardena, Lowe & Anderson (Gunawardena, Lowe, & Anderson, 1997) used content analysis to examine the construction of knowledge in an online debate. First, they examined existing techniques for identifying the negotiation of meaning and co-construction of knowledge in online learning environments. Their study determined that existing models were inadequate for understanding knowledge construction in computer mediated communication, but through their analysis, they developed a new model for examining online interactions for the co-construction of knowledge. Their approach is based on a constructivist theory of learning, which emphasizes that knowledge construction is evidenced by the introduction of new ideas, not the regurgitation of existing knowledge. Their model includes five progressive phases: Sharing/comparing, discovery & exploration, negotiation of meaning/co-construction of knowledge, testing and modification of proposed synthesis and agreement on constructed meaning. Marra, Moore & Klimczak (Marra, Moore, & Klimczak, 2004) compared Gunawardena et al's model for analyzing content with a model established by Newman, Web & Cochran (Newman, Webb, & Cochrane, 1996), determining that the Gunawardena et al (Gunawardena et al., 1997) model provides a more holistic view of discussion board flow & knowledge construction, although it requires the researcher to prepare a coding guide and procedures focused on a specific operationalization of knowledge construction in advance, as they did.

Our explication of the relationship between emergent grammars, social network structure and knowledge co-construction in CSCL contributes to the CSCL community's understanding of how CSCL log analysis can be used to identify meaningful relationships between performance and interaction. In Goggins, Laffey & Mascaro (Goggins, Laffey, & Mascaro, 2011) we show how event grammars emerge from micro-sequences of trace data in much the same way as conversational turns are revealed during conversation analysis of synchronous CSCL interactions. Relations between members are revealed through social network analysis of the same trace data. Event grammars are indexical, but smaller in scale and less ambitious in their reach than the process models Reimann (2009) proposes. Social network analysis brings the social nature of learning to the forefront, and helps to surface important relationships between CSCL participants during different periods of knowledge co-construction, interaction and performance referenced by other types of data. The integration of both types of log analysis with assessment of the degree of knowledge co-construction present in each contribution (post) to the CSCL system is a conceptual frame for the construction of more automated methods for analyzing the rich trace data available from CSCL systems. The contribution of this paper, then, is the description of a specific, progressive structure for capturing and analyzing CSCL traces which, if applied across technological systems, as we have begun to do with Sakai[1], Bugzilla[2], VMT[3] and Mylyn[4], will enable the automation of the more generalized analysis of interaction that is widely sought by leading CSCL researchers like Reimann, Suthers & Stahl.

Trace Data Resolution Levels

Our insight, derived from years of analysis and active participation in the field of CSCL suggests that how logs are structured at runtime, and for analysis, can represent the lived interactions of users through technology to greater and lesser degrees. In table one, we propose four data resolution levels for CSCL researchers to

[1] http://www.sakaiproject.org
[2] http://www.bugzilla.org
[3] http://www.mathforum.org/VMT
[4] http://tasktop.com/mylyn/

consider. The lowest level simply includes a user id, environment code, context id, session id, url, event type (read, post, etc), event object and a timestamp. Events are captured serially from the CANS system, and modified to support this analysis in each subsequent level. The expanding structure of capture and analysis described in table one is somewhat abstract, but reflects a clear description of how each progressive step provides the CSCL researcher with a more fine grained view of the experienced interactions of users. Access to the corresponding, generalized toolset will be made available as part of a conference presentation.

Table 1 - Data Resolution Levels for CANS Data Analysis

Data Resolution Level	**Data Resolution Level Description**
(1) Raw CANS Data	One event per row.
(2) Bi-Directional CANS Data	For example, if I read a discussion board topic that you created, then a connection is drawn between you and I. In addition to the data in "Raw CANS Data", this data set contains: 1. The distance in minutes between an event and the object (usually a discussion board) that the event is in response to. 2. An identifier for the object creator; this creates the social link.
(3) Exploded Bi-Directional CANS Data	Exploded Bi-Directional CANS data, as explained in Goggins, Laffey, Amelung & Gallager (2010b) and referenced in Goggins, Galyen and Laffey (2010a) recognizes the social form of online discussion. This includes recognition that when an individual participates in a discussion board or other interaction in a CSCL or other collaboration system, they frequently view more than one, specific post. This varies by environment. Effectively, one row is created for each artifact that is visible on a page when the page is viewed. In the systems we study, this is discerned from the timestamp, url and event object (artifact or discussion).
(4) Weighted Extraction of Exploded Bi-Directional CANS Data	The calculated time distance at level three can be transformed into a meaningful weighting factor for the analyst, depending on how other data gathered suggests weights should be calculated. This provides a concrete, automated method to support dependency graph construction with more refined weighting. For example, in many of our studies, interviews and field notes suggest a 3 to 4 day "cliffing" of the interaction weights is appropriate. These weighting "cliffs" depend on the environment studied.

Implications for Theory and Conclusion

The log capture and design techniques that we conceptualized in this paper provide a representative view socio-technical context in CSCL, and advance a broad range of CSCL research agendas. Data captured by most systems is designed for the convenience of system analysts and web based metrics, not for the analysis of social, collaborative behavior. CSCL log data should analytically useful, and it should represent social interactions that are both implicit and explicit. Time matters because more recent interactions are more salient for measuring the social nature of asynchronous interaction. Activity and context matter for connecting analytical dots, as Suthers, Reimann and Stahl all do. The analysis and capture we propose here takes a more full account of activity. For knowledge construction, performance and other questions in CSCL research, the full social context is critical. For our work, and the work of those in CSCL who have turned toward the social, the log analysis we propose here is a necessary and fundamental shift required when log analysis is a central method. We think our approach complements the recent methodological and analytical innovations of others in CSCL. The techniques and tools we have develop and describe here enable CSCL researchers to more fully incorporate social theories of learning (Bandura, 1977) into their work, because our approach makes the social visible.

REFERENCES

Amelung, C. (2007). Using Social Context and E-Learner Identity as a Framework for an E-Learning Notification System. *International Journal on E-Learning*, *6*(4), 501-517.

Bandura, A. (1977). *Social Learning Theory.* Englewood Cliffs, NJ: Prentice-Hall.

Bruckman, A. (2002). The Future of E-Learning Communities. *Communications of the ACM*, *45*(4), 60-64.

Cakir, M. P. (2007). *How online small groups create and use mathematical objects to do collaborative problem solving*. Proceedings from Proceedings of the 8th international conference on Computer supported collaborative learning.

Carroll, J. M., Rosson, M. B., Convertino, G., & Ganoe, C. H. (2006). Awareness and Teamwork in Computer Supported Collaborations. *Interacting With Computers*, *18*, 21-46.

Carroll, J. M., Neale, D. C., Isenhour, P. L., Rosson, M. B., & McCrickard, D. S. (2003). Notification and awareness: synchronizing task-oriented collaborative activity. *International Journal of Human-Computer Studies*, *58*, 605-632.

Dourish, P., & Button, G. (1996). Technomethodology: Paradoxes and Possibilities. *Proceedings of CHI96 Human Factors in Computing Systems,* 13-18.

Erickson, T., & Kellogg, W. A. (2000). Social Translucence: An Approach to Designing Systems that Support Social Processes. *ACM Transactions on Computer-Human Interaction*, *7*, 59-83.

Festinger, L. (1954). A Theory of Social Comparison Processes. *Human Relations*, *7*, 117-140.

Goggins, S., Laffey, J., & Mascaro, C. (2011). Context-Aware Event Log Analysis in CSCL: Emergent Grammars, Social Networks and Knowledge Construction. *Journal of Computer Mediate Communication*, *Under Review*.

Goggins, S., Galyen, K., & Laffey, J. (2010a). *Network Analysis of Trace Data for the Support of Group Work: Activity Patterns in a Completely Online Course*. Proceedings from ACM Group 2010, Sanibel Island, FL.

Goggins, S., Laffey, J., & Tsai, I.-C. (2007). *Cooperation and Groupness: Community Formation in Small online Collaborative Groups*. Proceedings from Proceedings of the ACM Group Conference 2007, Sanibel Island, FL.

Goggins, S. P., Laffey, J., Amelung, C., & Gallagher, M. (2010b). *Social Intelligence In Completely Online Groups*. Proceedings from IEEE International Conference on Social Computing, Minneapolis, MN.

Gunawardena, C. N., Lowe, C. A., & Anderson, T. (1997). Analyisis of Global Online Debate and the Development of an Interaction Analysis Model for Examining Social Construction of Knowledge in Computer Conferencing. *Journal of Educational Computing Research*, *17*(4).

Katz, N., Lazer, D., Arrow, H., & Contractor, N. (2004). Network Theory and Small Groups. *Small Group Research*, *35*, 307-332.

Laffey, J., Amelung, C., & Goggins, S. (2009). A Context Awareness System for Online Learning: Design Based Research. *International Journal on E-Learning*, *8*(3), 313-330.

Marra, R. M., Moore, J. L., & Klimczak, A. K. (2004). Content Analysis of Online Discussion Forums: A Comparative Analysis of Protocols. *Educational Technology Research and Development*, *52*(2), 23 - 40.

Newman, D. R., Webb, B., & Cochrane, C. (1996). A Content Analysis Method to Measure Critical Thinking in Face-to-Face and Computer Supported Group Learning. *Interpersonal Computing & Technology*, *3*, 56-77.

Olson, G. M., Herbsleb, J. D., & Rueter, H. H. (1994). Characterizing the sequential structure of interactive behaviors through statistical and grammatical techniques. *Human-Computer Interaction*, *9*(4), 427-472.

Reimann, P. (2009). Time is precious: Variable-and event-centred approaches to process analysis in CSCL research. *International Journal of Computer-Supported Collaborative Learning*, *4*(3), 239-257. doi:10.1007/s11412-009-9070-z

Reimann, P., Frerejean, J., & Thompson, K. (2009). *Using Process Mining to Identify Models of Group Decision Making in Chat Data*. Proceedings from CSCL 2009, Rhodes, Greece.

Stahl, G. (2009a). *For a science of group interaction*. Proceedings from Proceedings of the ACM 2009 international conference on Supporting group work.

Stahl, G., Ou, J. X., Cakir, M. P., Weimar, S., & Goggins, S. (2010). *Multi-User Support for Virtual Geogebra Teams*. Proceedings from GeoGebra NA2010, Ithica,NY.

Stahl, G. (2006). *Group Cognition: Computer Support for Building Collaborative Knowledge.* Boston, MA: MIT Press.

Stahl, G. (2009b). Toward a Science of Group Cognition. Retrieved July 2, 2009 from http://www.ischool.drexel.edu/faculty/gerry/vmt/book/28.pdf

Suthers, D., Dwyer, N., Medina, R., & Vatrapu, R. (2007). *A framework for eclectic analysis of collaborative interaction*. Proceedings from Proceedings of the 8th international conference on Computer supported collaborative learning.

Suthers, D. D. (2006). Technology affordances for intersubjective meaning making: A research agenda for CSCL. *International Journal of Computer-Supported Collaborative Learning*, *1*(3), 315-337. doi:10.1007/s11412-006-9660-y

Suthers, D. D., Dwyer, N., Medina, R., & Vatrapu, R. (2010). A framework for conceptualizing, representing, and analyzing distributed interaction. *International Journal of Computer-Supported Collaborative Learning*, *5*(1), 5-42. doi:10.1007/s11412-009-9081-9

Wasserman, S., & Faust, K. (1994). *Social Network Analysis: Methods and Applications.* New York, NY: Cambridge University Press.

Wenger, E. (1998). *Communities of Practice: Learning, Meaning and Identity.* New York: Cambridge University Press.