Beyond Threaded Discussion: Representational Guidance in Asynchronous Collaborative Learning Environments

D. Suthers*, R. Vatrapu, R. Medina, S. Joseph, N. Dwyer

Department of Information and Computer Sciences, University of Hawai`i at Manoa, 1680 East West Road POST 309, Honolulu, HI, 96822, USA

Abstract

Although most online learning environments are predominately text based, researchers have argued that representational support for the conceptual structure of a problem would address problems of coherence and convergence that have been shown to be associated with threaded discussions and more effectively support collaborative knowledge construction. The study described in this paper sets out to investigate the merits of knowledge mapping representations as an adjunct to or replacement for threaded discussion in problem solving by asynchronously communicating dyads. Results show that users of knowledge maps created more hypotheses earlier in the experimental sessions and elaborated on them more than users of threaded discussions. Participants using knowledge maps were more likely to converge on the same conclusion and scored significantly higher on post-test questions that required integration of information distributed across dyads in a hidden profile design, suggesting that there was greater collaboration during the session. These results were most consistent when a knowledge map with embedded notes was the primary means of interaction rather than when it augmented a threaded discussion.

The paper also offers a methodological contribution: a paradigm for practical experimental study of asynchronous collaboration. It is crucial to understand how to support collaborative knowledge construction in the asynchronous settings prevalent in online learning, yet prior experimental research has focused on face-to-face and synchronous collaboration due to the pragmatic problems of conducting controlled studies of asynchronous interaction. A protocol is outlined that enables study of asynchronous collaboration in a controlled setting.

Journal keywords: computer-mediated communication; cooperative/collaborative learning; human-computer interface

Additional keywords: knowledge maps, representational guidance

---

*Corresponding Author: Dan Suthers, Department of Information and Computer Science, University of Hawai`i, 1680 East West Road, POST 309, Honolulu, HI 96822, USA. Tel: 1.808.956.3890, Fax: 1.808.956.3548.
Email: suthers@hawaii.edu

Introduction

The use of electronic media for online learning has expanded greatly in the past decade (Allen & Seaman, 2005; Mayadas, 1997), yet too often implementations use pre-existing Internet technology to “deliver” conventional but ineffective pedagogical approaches, rather than adopting or inventing new technologies specifically designed to support effective approaches to learning (Hill, Wiley, Nelson, & Han, 2003). Research on learning and instruction has shown the importance of learners’ active participation in expressing, testing, and revising their own knowledge (e.g., Chi, Bassok, Lewis, Reimann, & Glaser, 1989; Perkins, Crismond, Simmons, & Unger, 1997; Scardamalia & Bereiter, 1994). Therefore, electronic media should support such engagement, leveraging the computational medium’s strengths for education: its representational and analytic capabilities, its interactivity, and its networking support for collaboration. Prior research also shows the importance of representational aids, such as dynamic notations, knowledge maps, and simulations for individual problem solving (e.g., Kotovsky & Simon, 1990; Larkin & Simon, 1987; Novick & Hmelo, 1994; Zhang, 1997) and learning (e.g., Koedinger, 1991; Novak, 1990; Scardamalia et al., 1992; Snir, Smith, & Grosslight, 1995)); and of social processes such as collaboration and mentoring to learning (e.g., Brown & Campione, 1994; Hiltz, Coppola, Rotter, & Turoff, 2000; Lave & Wenger, 1991; Scardamalia & Bereiter, 1991; Webb & Palincsar, 1996). Representational and social scaffolding can be combined to provide representational support for collaborative learning (Dillenbourg, 2005; Roschelle, 1996; Suthers & Hundhausen, 2003), a form of synergistic scaffolding (Tabak, 2004). Research on representations that are constructed by learners during collaboration (Suthers & Hundhausen, 2003) and representations used as a discourse medium (Baker & Lund, 1996; Guzdial & Hmelo, 1997) has shown that the choice of representation can lead to different forms of learning discourse.

A separate but related line of research on computer-mediated communication (CMC) has identified several problems related to typical discourse representations through which people communicate online (e.g, threaded discussion and chat). Although the reviewability and revisability (Clark & Brennan, 1991) of contributions to discussion forums may support more reflective contributions (Hawkes & Romiszowski, 2001), facilitate participation of those who hesitate to participate in spoken discussions (Heckman & Annabi, 2003) and enable forms of participation that do not take place in the classroom (Kitade, 2000), there are also disadvantages related to the extra effort required for written communication and engaging students’ participation (Frank, 2006). Issues sensitive to the design of the medium itself include *incoherence* due to the violation of adjacency conventions for topic maintenance (Herring, 1999) and the coarse granularity of referencing (Reyes & Tchounikine, 2003); and *lack of convergence* due for example to the intrinsically divergent representations used in threaded discussion (Hewitt, 2001) and a bias towards addressing recently posted messages (Hewitt, 2003). The shared agreement or knowledge being constructed through discourse is not made explicit by typical CMC tools, and hence it is difficult to find relevant contributions, place one’s own contribution in the relevant context, or quickly assess the outcome of the discourse (Suthers, 2001; Turoff, Hiltz, Bieber, Fjermestad, & Rana, 1999). The fundamental problems are a lack of integration of discourse representations with other representations and a lack of explicit construction of the desired outcome of the collaboration, leading to weak support for online collaborative knowledge construction. In response to these problems, Suthers (2001) proposed better online support for *artifact-centered discourse* (discourse that makes reference to and is tightly integrated with visual or textual artifacts), and suggested that synergistic benefits may be obtained if these artifacts are also knowledge representations being constructed by the learners. Specifically, if each contribution to the discourse can be referenced to a component of the knowledge representation, coherence may improve because the conceptual relevance of each contribution is clear (Mühlpfordt & Wessner, 2005; van der Pol, Admiraal, & Simons, 2006), and convergence may improve because multiple contributions referencing a given topic are collected together. The knowledge representation can also serve as a summary of the status of the collaboration, available to learners and mentors to support reflection and assessment.

In summary, prior work suggests the potential value of representational guidance for social processes of learning, and specifically of the potential value of collaborative media in which participants can make conceptual structure explicit. The present study constitutes an experimental test of these ideas. Participants were enabled to construct an explicit representation of the topics and conclusions of the discourse itself as they interacted. Two forms of artifact-centered discourse were compared to each other and to a threaded discussion control condition. Since our interest is online collaborative learning, which commonly includes a strong asynchronous component (Mayadas, 1997), we confronted the problem of experimentally studying asynchronous collaboration. A pragmatically viable approach to empirical evaluation of asynchronous collaboration is an additional contribution of this paper. The study was also designed to provide rich data for further analysis of how participants appropriate media affordances for collaborative knowledge construction, a line of analysis to be reported in other publications and that, as it turns out, is motivated by the quantitative results reported in this paper.

The paper is organized as follows. First we specify our research hypotheses and explain how these are reflected in the software designs that define the experimental treatments. The protocol for experimental study of asynchronous collaboration is described in its own section, as this bridges from the software design to the methodology and is of interest in its own right. The remaining sections follow the traditional sequence of methods, results and discussion.

1. Hypotheses and software designs

The scientific objective of this work is to understand the role of representational tools in collaborative knowledge construction processes, and the associated engineering objective of this work is to improve online collaborative knowledge construction environments. The following hypotheses capture the relationship between the scientific and engineering objectives by motivating the software designs that constitute the treatment conditions.

1.1 Hypotheses

From the standpoint of constructivism (Von Glasersfeld, 1995), *knowledge construction* goes beyond an information transfer conception of learning by placing the responsibility for the creation of knowledge on the learner. Knowledge construction seeks systematicity, coherence, and convergence as participants engage in meaning-making to extend their understanding (Wells, 1999). Knowledge construction is *elaborative*, because understanding is improved when the implications of an idea are explored; *integrative*, because coherence is improved when connections are formed between distinct elements of one's understanding; and *reflective*, because one must be aware of and assess the state of one's own knowledge to determine where improvements can be sought, and in particular in order to identify opportunities for elaboration and integration. *Collaborative* knowledge construction admits the possibility that these processes can take place in joint as well as individual acts of meaning-making (Stahl, 2006; Suthers, 2006b). The first hypothesis claims that conceptually oriented representations facilitate these elaborative, integrative and reflective processes in groups because explicit representations of collective understanding in a persistent, shared and inspectable medium enable these processes to be distributed across individuals (Hutchins, 1995).

H1: Collaborative knowledge construction is more effectively supported by environments that make conceptual objects and relations explicit.

The hypothesis specifically privileges conceptual representations, not just any representation in a persistent shared medium. Communication media that are structured by discourse relations such as reply structure (e.g., threaded discussions) capture the historical development of discourse rather than its conceptual content, making it difficult to make contributions that move it forward (Suthers, 2001; Turoff et al., 1999). Explicit representations of conceptual structure have the advantages that they encourage participants to clarify their thinking (Brna, Cox, & Good, 2001), make this thinking visible to others (Bell, Davis, & Linn, 1995), provide resources for subsequent conversation (Roschelle, 1996; Scardamalia, 2004), can guide students' argumentation to include disconfirming as well as confirming evidence (Toth, Suthers, & Lesgold, 2002; Veerman, 2003), and can function as a "convergence artefact" that expresses the group's emerging consensus (Hewitt, 2001; Suthers, 2001).

This first hypothesis does not specify the relationship between the knowledge representations and the discourse that accompanies the creation of those representations. The next two hypotheses are alternative elaborations of H1, arguing for either maintaining the distinction between discourse and knowledge representations or combining the two. We begin with a strong version of H2 that will be modified later in this section:

H2': Collaborative knowledge construction is more effectively supported if there is no representational distinction between discourse and knowledge.

This hypothesis is motivated by the observation that knowledge and discourse are tightly related in practice. Knowledge lives in interaction: it is not possible to separate them (Garfinkel, 1967); therefore one might conclude that tools for collaboration should not attempt to separate them. According to this view, it is not possible to dichotomize our interactions by saying "that is discussion" and "that is the knowledge that is the product of the discussion." For example, contributions in the discussion might be reinterpreted, elaborated, and brought to bear on other situations in a manner that constitutes them as shared knowledge. This argument for H2' states that since discourse and knowledge cannot be distinguished, the representational medium should not force this distinction, but should instead provide a collection of semiotic resources with and through which participants can interact in a mutual construction of knowledge.

One could argue that a literal translation of the nature of knowledge to a recommendation for the design of tools for collaboration is a category mistake, confusing *knowledge* with *conceptual representations*. An argument about the nature of knowledge need not necessarily be literally mirrored in the representational resources we provide. Even if designers provide separate "knowledge" and "discourse" representations, users may not respect this distinction. Collaborators will distribute their interaction across all mutable media (Suthers, Hundhausen, & Girardeau, 2003), and knowledge may yet live in interaction regardless of how this interaction is distributed across representational media chosen by participants to meet their needs.

Yet, a more pragmatic case for an integrated representation can be made. One could argue that discourse representations should be embedded in or mixed with the conceptual representations to contextualize the discussion and facilitate ease of reference (e.g., by simple attachment of notes to the objects to which they refer). Suthers (2001) called this *embedded artifact-centered discourse* because the discourse is embedded in the artifact under discussion. A usability argument can also be made: it may be easier to manage a single workspace than interactions distributed across multiple tools. This pragmatic justification is reflected in a weaker version of H2':

H2: Collaborative knowledge construction is more effectively supported if discourse and conceptual representations are tightly integrated.

The third hypothesis is motivated by the observation that discourse structures and conceptual structures are different: discourse relies on regularities in adjacency and focus shifts for coherence (Grosz & Sidner, 1986; Sacks, Schegloff, & Jefferson, 1974), while conceptualizations may be organized according to diverse ways of modeling or systematizing knowledge about the world. Therefore, separate tools will enable designers to optimize representations to meet the distinct structural needs of discourse and conceptualization in a given domain of discourse. This point leads us to the third hypothesis, which is in direct opposition to the second:

H3: Collaborative knowledge construction is more effectively supported if the distinction between discourse and conceptual is reflected in the representations provided.

In *linked artifact-centered discourse* (Suthers, 2001), discourse media such as threaded discussions are maintained separately from knowledge representations or other disciplinary representations being discussed, but referential links can be made to the relevant parts of the latter representations. A linked approach attempts to maintain one major advantage of the embedded approach, the contextualization of contributions, while addressing deficiencies and adding other advantages. When given its own representation, the chronological and reply structures of the discourse may be maintained, and discussions that generalize across multiple objects in the representations are more natural. Yet, explicit "linking" or reference of discourse contributions to conceptual objects resolves some of the incoherence resulting from the violation of contiguity of related discourse contributions that is so common in electronic media. Explicit referencing has been explored in software implementations by others (Mühlpfordt & Wessner, 2005) as well as our own.

A synthetic view is offered by Hoadley and Enyedy (1999), who call for filling the "middle space between communication and information interfaces." Discourse and knowledge representations are both valued, but support is required to help bridge between the "dialogical" and "monological" forms of learning that they support as knowledge moves between social and individual realms (Vygotsky, 1978). The present work is an exploration of this middle ground.

1.2 Software environments

We constructed three software environments (Figures 1-3) in order to test these hypotheses. All three of the environments have an "information viewer" on the left in which materials relevant to the problem are displayed. This information viewer functions as a simple web browser, but presentation of materials is constrained as discussed in the next section.

All three environments have a shared workspace or "information organizer" on the right hand side in which participants can share and organize information they gather from the problem materials as well as their own interpretations and other ideas. The three environments differ on the nature of the "information organizer," as described below. Changes made to the workspace by each participant are propagated to other participant's displays of the same workspace under a protocol to be discussed in the next section. In all three environments, mutual awareness of participants' activity is also supported as follows: yellow circles are used to mark information posted by the user of the environment but not yet opened by his or her partner, while red triangles are used to mark new information from the user's partner that the user has not yet opened. All contributions must be opened in order to be read.

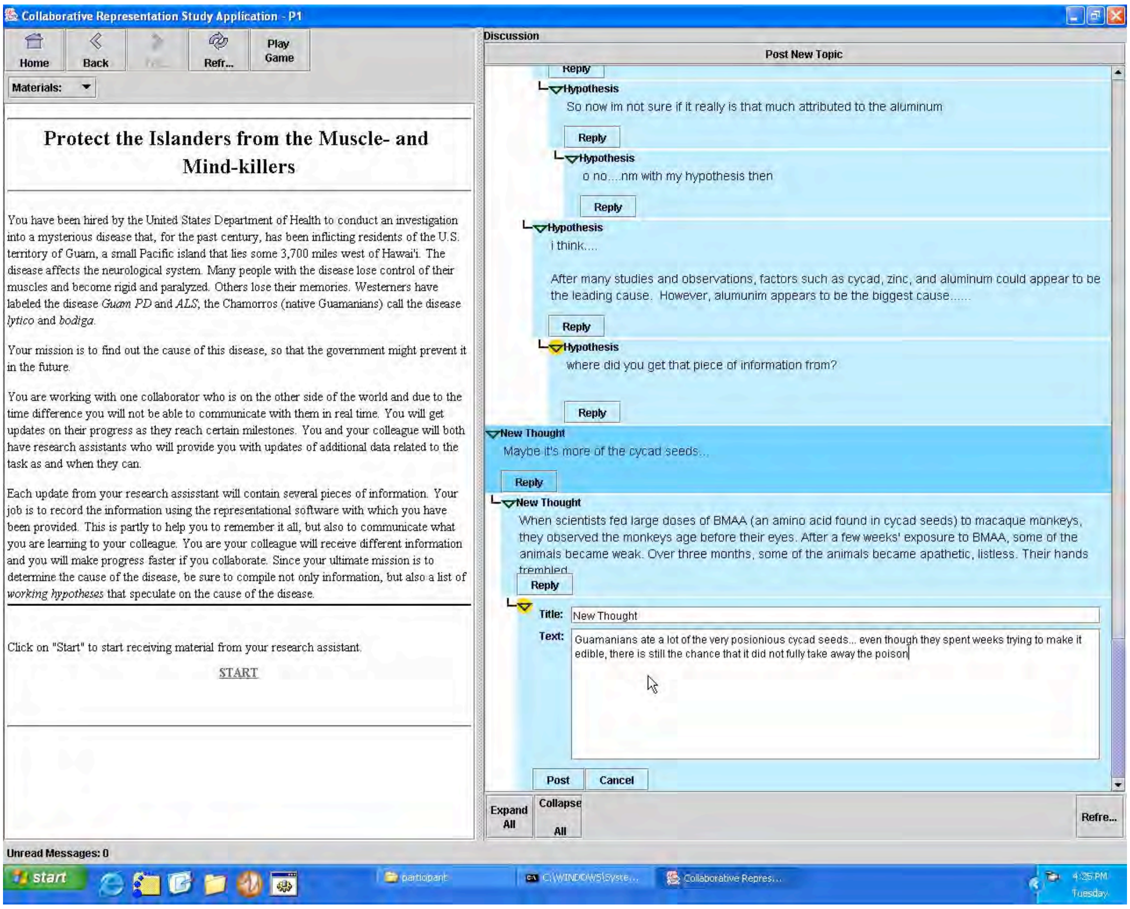

Figure 1. "Text" environment (threaded discussion)

1.2.1 Text condition

The shared workspace in the "Text-only environment," or "Text" condition for short, is a conventional threaded discussion tool (Figure 1). This environment functions as the control condition for testing the above hypotheses, since the workspace only provides explicit support for representation of discussion structure (subject headings and reply relations).

1.2.2 Graph condition

The shared workspace in the "Graph-only environment," or "Graph" condition for short, consists of an integrated node-and link graphing tool derived from Belvedere (Suthers et al., 2001) in which one can visually express relations of evidence between data and hypothesis objects and also post notes that can be free-floating or attached to specific objects. Such graphs fall within the category of tools called "knowledge maps" (a generalization of concept mapping, Novak, 1990). In this paper, we use "evidence map" to refer to the specific representational tool used in the study, and "knowledge map" to refer to the category of conceptually explicit representations.

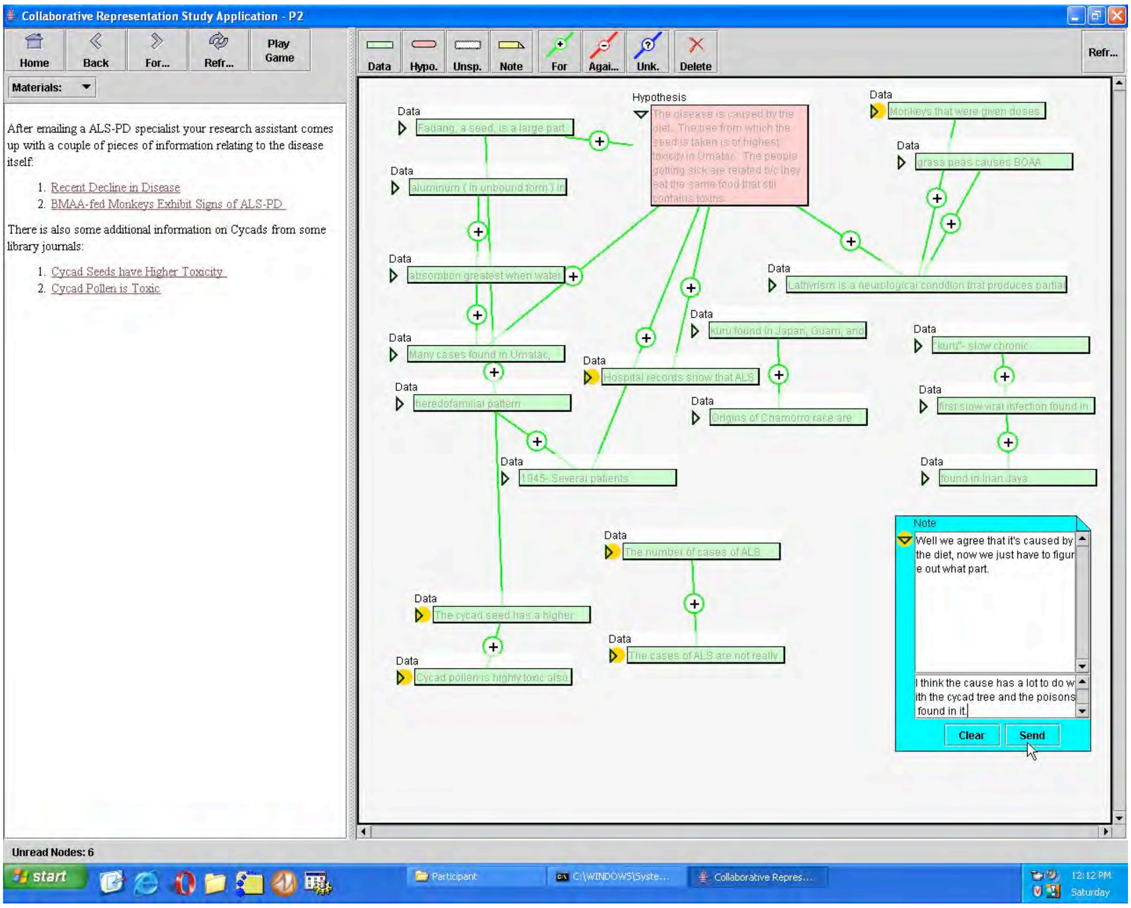

Figure 2. "Graph" environment (knowledge map)

Like the Mixed environment (described next), the Graph workspace includes tools for constructing conceptual objects under a simple typology relevant to the task of reasoning about evidence, including *data* (green rectangles, for empirical information) and *hypotheses* (pink rectangles, for postulated causes or other ideas). There are also linking tools for constructing *consistency* ("for") and *inconsistency* ("against") relations between other objects, visualized as green links labeled "+" and red links labeled "-" respectively. "Unspecified" objects and "unknown" links are also provided for flexibility. Finally, a *note* object (shown in Figure 2) supports a simple linear (unthreaded) discussion that appears similar to a chat tool, except that a note is interactionally asynchronous and one can embed multiple notes in an evidence map and link them like any other object.

Our Graph workspace reflects the weaker hypothesis H2, which claims that discussion is best supported in a contextualized manner, embedded in the conceptual representation for ease of reference. This configuration also has the advantage of simplicity over the Mixed condition (described next) in the sense that there is just one workspace. The stronger hypothesis H2', that there should be no representational distinction between knowledge and discourse, is insufficiently reflected in the Graph software because the presence of notes that support discussions among conceptual representations still dichotomizes the elements of interaction. Another line of work initiated in (Dwyer & Suthers, 2005, in press) is addressing the problem of how to provide more flexible media for interaction through representations.

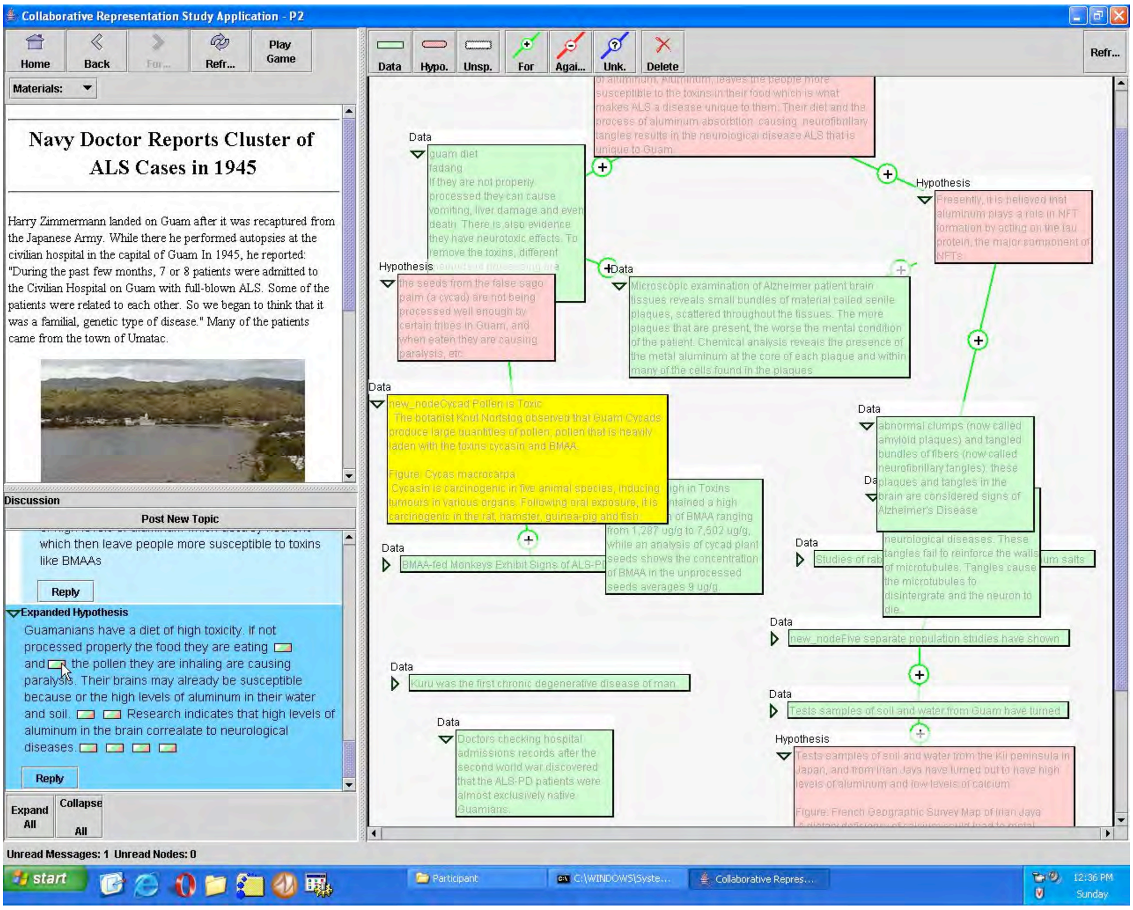

Figure 3. "Mixed" environment (threaded discussion linked to knowledge map)

1.2.3 Mixed condition

The shared workspace of the "Mixed" condition includes both a threaded discussion tool and an evidence mapping tool for representing conceptual structure in the same manner as the Graph condition, except that there are no embedded notes in the Mixed version of the evidence map. Instead, one can embed references to evidence map objects in the threaded discussion messages simply by clicking on the relevant graph object while composing the message. The references show up as small icons in the message (Figure 3). When the reader selects the icon, the corresponding object in the evidence map will be highlighted, indicating the intended referent (as shown in Figure 3). This environment is motivated by H3, which claims that separate representations are needed to optimize discussion and conceptual representations, but that they should be logically "linked" for referential purposes.

1.3 Experimental design

H1 is tested by comparing performance of users of the Text environment to performance of users of the Graph and Mixed environments. There are two ways in which this comparison could be done. Given that H1 predicts that the presence of a conceptual representation will be beneficial, we could compare the performance of Text users to the performance of Graph and Mixed users in aggregate. However, there are many choices to be made in designing software environments, and we anticipated that the implementation chosen could obscure the viability of H1. Specifically, we have two competing hypotheses concerning the best implementation strategy, H2 and H3, which motivate the Graph and Mixed implementations, respectively.

Therefore, in order to determine whether *some* implementation of a conceptual representation is better than threaded discussion alone, we planned to test H1 through two sets of comparisons: Text versus Graph and Text versus Mixed. The competing hypotheses H2 and H3 are tested by comparisons of the Graph and Mixed conditions to each other.

"Performance" includes process and outcome measures. Planned comparisons on process measures included the number of hypotheses proposed and the extent to which these hypotheses were elaborated on or integrated with evidence. Planned comparisons on outcome measures included the quality of conclusions reached, convergence of participants on the same conclusion, the extent to which participants relied on shared information for their essays, individual memory for different kinds of information, and usability evaluation of the software. Usability measures were included to address the possibility that the introduction of conceptual representations or other specific design choices could make the software more complex in ways that impact upon the hypotheses. These measures will be described in further detail later. In addition to these specific measures, we also recorded sufficient data to enable exploratory analysis of the experimental sessions to uncover other unanticipated differences between treatment groups that might bear upon our hypotheses.

Our protocol for laboratory simulation of asynchronous communication is described next, as it underlies much of the methods and may be valuable to others as a research strategy in its own right. Subsequently we will return to other aspects of the methods, results and discussion.

2. A protocol for experimental study of quasi-asynchronous collaboration

At the outset of this work several years ago, the majority of controlled experimental studies of computer-mediated communication had been undertaken in synchronous collaboration settings. We viewed this as problematic because a significant portion of applications of computer-mediated communication to online learning are primarily asynchronous. Based on personal communication with other researchers, the first author concluded that a major reason for the lack of studies of asynchronous collaboration is logistical: it is easier to conduct a study in which participants come to the laboratory for one experimental session rather than a study in which participants must return to the controlled setting at different times repeatedly over a period of time. In the latter situation, the experimenter must be concerned with a potentially higher attrition rate (a significant amount of work can become useless if a participant fails to show up for the final experimental session), and with whether participants would engage in other activities between experimental sessions that invalidate the assumptions of a controlled design. Faced with these challenges, we designed a study protocol that simulates many of the properties of asynchronous communication while still enabling us to conduct experimental sessions with participants in the laboratory at the same time.

The fundamental criterion was that there be no particular timing constraint between the actions of participants (e.g., waiting for the participant's action before being able to continue one's own work), nor temporal affordances to be exploited in a synchronous manner (e.g., sending a message and expecting an immediate reply). A second aspect of asynchronous work that we sought to simulate (albeit necessarily less faithfully in the laboratory setting) is that one might stop working on a problem for a while, do something else, and then return to the work. We achieved these desiderata through a study protocol in which (1) participants took occasional "breaks" from their work to play a computer game, and (2) the work of the other participant became available only after these breaks. We discuss this protocol in further detail below.

2.1 Presentation of materials and session breaks

The study materials were divided into twelve (12) sets of materials, each set consisting of a brief introduction and links to four pages, each page containing a short article. The contents of these articles will be described later. Each of two participants was assigned six sets of materials, presented in six "study sessions" (to be distinguished from the experimental session as a whole). In each study session, participants were expected to read the four articles and update the shared workspace, as they deemed appropriate. When done, they could not obtain the next set of materials until they had "taken a break" by playing a computer game, a Java version of Tetris™, for 1-5 minutes (pilot studies showed that longer breaks made the experiment excessively long). Tetris™ was chosen for its familiarity and because it presents a perceptual motor activity quite different from the cognitive task of the study, in this sense constituting a break from the primary task. Paper-based activities were also considered, but rejected in favour of automating the timing and logging of breaks.

2.2 Protocol for workspace updates

The actions of each participant in the shared workspace were not displayed immediately in the other participant's workspace. As a person worked, the actions of that person were sent to the other participant's client application, but were queued rather than displayed. When a participant started a new study session with the next set of articles, all of the currently queued actions on that client were displayed. Conflicts that might arise when both participants edited the same object were resolved through *operational transformations* (Sun, Jia, Zhang, & Yang, 1997). Operations O_a enacted by client A and O_b enacted by client B are transformed into O_a' and O_b' such that $O_a'(O_b) = O_b'(O_a)$. In other words, if A's transformed operations O_a' are applied to B's workspace after B has made changes O_b, then the result will be the same as when B's transformed operations O_b' are applied to A's workspace after A has made changes O_a. As a result, clients A and B are guaranteed to converge on the same state. The delayed updating protocol simulates one aspect of the experience of asynchronous collaboration: a participant sees what one's partner has done upon returning to a workspace after a period of time. It excludes the possibility of synchronous conversation in which one participant posts a message in the workspace and receives an immediate reply.

Pilot studies suggested that this protocol as just stated would be too strict. One participant would sometimes fall far behind another, who was wondering whether any work was being done in the workspace. Also, we recognized that in an asynchronous environment sometimes two people are working at the same time, and it is possible to get updates by refreshing the workspace with respect to a server. This was not the case with the pilot-tested software as there was no way to receive an update of the partner's work without taking a Tetris™ break. To address these concerns we cautiously introduced a "refresh" feature that enables one to get all updates to that point in time. We were concerned that, upon discovering this feature, participants might use it to engage in synchronous interaction by alternating between posting messages and refreshing the workspace while waiting for a reply. However, in our second set of pilot studies and in the actual study itself, this did not happen very often. Participants used the refresh feature primarily at the end of the study session when one person finished first and was waiting for the other person to finish their work. (The instructions required that they come to a final conclusion based on material they had shared with each other.)

Table 1. Grounding constraints in ALN and QAP interactions

	ALN	QAP
Copresence: *A and B share the same physical environment.*	False	False
Visibility: *A and B are visible to one another.*	False	False
Audibility: *A and B communicate by speaking.*	False	False
Contemporality: *B receives at roughly the same time as A produces.*	False	False
Simultaneity: *A and B can send and receive simultaneously.*	False	False
Sequentiality: *A's and B's turns cannot get out of sequence.*	False	False
Reviewability: *B can re-view A's messages.*	True	True
Revisability: *A can revise message for B (edit before sending)*	True	True

In order to assess the extent to which this protocol simulates asynchronous interaction, we compared our protocol to typical asynchronous online learning scenarios (Mayadas, 1997). The comparison uses Clark and Brennan's "grounding constraints" (Clark & Brennan, 1991), well known dimensions for analyzing the affordances of communication media (called "constraints" because they offer further information to constrain interpretations, thereby facilitating convergence on "common ground"). The comparison shows that asynchronous interaction as enacted in asynchronous learning networks (ALN) and our quasi-asynchronous protocol (QAP) provide or fail to provide exactly the same grounding constraints, and are therefore equivalent according to these constraints (Table 1).

We are not naïve enough to claim that the quasi-asynchronous protocol produces a situation literally identical to ALN. It is interesting that Clark's constraints do not capture the ways in which our study protocol differs from "real" asynchronous learning. These differences include the time-span of interaction (possibly spread over days in ALN rather than a few hours, providing time to think about a problem between sessions), the knowledge in QAP that one's partner is present in the same building working on the same problem at the same time (which may influence participants even though they cannot take advantage of this communicatively), and the socio-cultural situation (participant in an experiment versus student in an online course). However, the media affordances being studied are identical.

3. Methods

In this section we summarize the remaining aspects of experimental method. Further details necessary to replicate this work or conduct similar studies are presented in the appendices.

3.1 Participants

Pairs of participants were recruited from introductory courses in the College of Natural Sciences at the University of Hawai`i. Participants were paid US$50 each for participating in the study. We recruited participants in pairs of acquaintances so as to eliminate the social awkwardness of interaction between persons who do not know each other (found to be problematic in our previous work) and to provide a more realistic model of collaborative discovery (Okada & Simon, 1997).

Excluding pilot studies, we conducted a total of 30 experimental sessions involving 30 pairs or 60 participants. There were 10 pairs of participants (20 participants) for each of three treatment groups: Text, Graph and Mixed. Conditions were gender-balanced because previous studies showed that gender pairing substantially influenced the style of interaction. Each treatment group included 4 female-female, 4 female-male and 2 male-male dyads. We verified that there were no statistically significant differences between the three treatment groups on age and grade point average. A two-way ANOVA of the three treatment conditions (Graph, Mixed and Text) and the two dyad roles (Participant 1 and Participant 2) with respect to age was not significant ($F(2, 54) = 0.18$, $p = 0.8361$). Similarly, a two-way ANOVA with respect to official university records of their cumulative grade point average (CGPA) was also not significant ($F(2, 54) = 1.20$, $p = 0.3105$). We also verified through a pre-experiment questionnaire that none of the participants had prior experience with the study problem.

3.2 Materials

3.2.1 Topics

The study presented participants with "science challenge" problems, consisting of issues in science and public health. The "Riddle of the Time Traveling Iguanas" problem (resolving a discrepancy in the dating of speciation of Galapagos iguanas) was used as a "warm-up" exercise in which participants could become familiar with the software and with collaborating through that software. The "Protect the Islanders from the Muscle- and Mind-killers" problem challenged participants to identify the cause of a disease on the island of Guam known as ALS-PD. This disease has been under investigation for over 50 years, in part because it shares symptoms with Alzheimer's and Parkinson's diseases (Lieberman, 2004). Over the years numerous diverse hypotheses have been proposed and an even greater diversity of evidence of varying types and quality explored. Only recently have investigators converged on both a plausible disease agent (a neurotoxic amino acid in the seed of the Cycad tree) and the vector for introduction of that agent into people (native Guamians' consumption of fruit bats that eat the seed). These facts along with the relative obscurity of the problem make it a good problem to use when one wants participants to grapple with interpretation of multiple explanations and ambiguous data. All participants began with a mission statement that provided the problem description and task information. This mission statement is displayed in the left hand side of Figure 1 and is also available online.[1]

3.2.2 Information Articles

Source materials were provided in the form of short articles, typically consisting of one to two brief paragraphs and an image or two. An example article is shown in the left hand side of Figure 3. A complete list of source materials for the ALS-PD problem is available online.[2] Each article was designed to provide one key item of information relevant to the generation or evaluation of a hypothesis. The remaining information in a given article elaborated on this item or provided tangentially related "distractor" information. We designed the articles to provide evidence both for and against five major hypotheses (letters indicate codes used in Table 2): (A) aluminium levels in water and soil, (G) genetic causes, (Z) zinc levels in water, (C) consumption of cycad flour, and (B) consumption of fruit-eating bats as a source of the cycad toxin. The articles also included a mission statement and other general information about the disease and its demographics (D).

The information needed to draw a conclusion about any given hypothesis was distributed across more than one article. The articles presented to a given participant were clustered into six groups of four articles (for example, as shown in the left hand side of Figure 2). Each group of articles was presented to the participant in a "study session" separated by games of

[1] http://lilt.ics.hawaii.edu/lilt/papers/2006/Suthers-et-al-CE-2006/
[2] ibid

Tetris™ as described previously. Each participant in a dyad received a different sequence of articles, although there was some overlap between both the articles given to participants and the information in non-identical articles, as described below.

Table 2: Distribution and sequencing of information articles across participants and sessions.

Session#	P1's Articles				P2's Articles			
1	A7+	G3-	A1+	A2+	G1+	G2+	C1+	C2+
2	D1	D4	A3+	A5+	D6	C3+	C7+	C8+
3	C1+	B2+	A6+	D2	B1+	B5+	A2-	A1-
4	C6+	D5	C3-	G1-	A3-	Z1+	C5+	Z2+
5	Z1-	G2-	C2-	D3	C10+	C9+	A4-	B4+
6	C5-	B3+	A4+	C4-	C4+	C11+	C1-	Z3+

Key: The articles provide background information about the disease and its demographics (D) and evidence both for (+) and against (-) five major hypotheses:
A: Aluminum levels in water and soil
B: Consumption of bats as a vector for the cycad toxin
C: Consumption of cycad flour as source of cycad toxin
G: Genetic causes
Z: Zinc levels in drinking water

3.2.3 Distribution of Evidence Across Participants

We used a "hidden profile": a classic paradigm in studies of group problem solving (Stasser & Stewart, 1992) in which information is distributed across participants such that a participant relying only on information he or she directly received would come to a suboptimal conclusion. Table 2 shows the complete distribution of materials across the participants. For example, one participant (P1) received evidence for aluminium as a disease agent (A1+ through A7+ in Table 2) and evidence against genetic causes (G1- through G3-), while the other participant (P2) received evidence for genetic causes (G1+ and G2+) and evidence against aluminium (A1- through A4-). Information sharing between participants was required in order for either participant to reject these and other hypotheses and identify the most complex explanation that incorporates the evidence implicating a toxin derived from cycad seeds (C1+ through C11+), but addresses the low toxicity of prepared cycad flour (C2- and C4-) by identifying bats as an alternative vector via which the toxin enters humans (B1+ through B5+). Because of this distribution of information, we can draw conclusions concerning information sharing by eliciting participants' beliefs and evidence for those beliefs at the end of the experimental session.

3.2.4 Sequencing of Articles

The sequencing of articles was designed to require integration over time, rewarding use of the information organizing workspace to make relevant information available in later study sessions. This sequencing is shown in Table 2. For example, one line of evidence for the bats as vector requires a cross-session connection between the recent decline in the disease (D6, presented to P2 in study session 2) and that of the bats decline (B3+, presented to P1 in study session 6). Further evidence requires cross-session connections between the prevalence of the disease in male patients (D3, presented to P1 in study session 5), dining practices in which primarily men eat bat skins (B5+, presented to P2 in study session 3), and information on the higher toxicity of bat skins (B4+ presented to P2 in study session 5).

By the end of the last study session, the evidence is ambiguous for all hypotheses, but a combination of the cycad as a source of the toxin and bats as a vector for introduction of that toxin to humans incorporates the well developed case for cycads while addressing the evidence against cycad flour toxicity. Due to the distribution of conflicting evidence, sharing of information across participants and study sessions is needed to expose the weaknesses of the aluminum, genetics and zinc hypotheses, as well as to construct the more complex explanation involving bats and cycad seeds.

3.3 Procedure

After signing of consent forms, participants filled out a demographic survey. They were then introduced to the software and format of the study sessions through a standardized set of instructions and demonstrations designed to be as equivalent as possible across all conditions. The instructions included:

1. A brief tutorial and demo on how to use the computer software.
2. Instructions to practice with the software on the first warm-up problem
3. Instructions for the main study problem.
4. Information about the short essay and questionnaire to be completed after finishing the main study problem.
5. Information about a brief online test to be given one week later.

Participants were then led to their respective stations in different rooms from each other, and began work on a “warm-up” problem, speciation of the Galapagos iguanas, to familiarize themselves with the software. After a maximum of 30 minutes of work on the warm-up problem, participants were instructed to halt work and begin work on the main problem; Guam ALS-PD. Participants were given up to 120 minutes to work through all of the information available for this problem. The update protocol described earlier was applied during the experiment to present the six “study sessions” of materials that we have just reviewed.

At the conclusion of their final study session, each participant working alone was given up to 30 minutes to write an essay on the hypotheses that were considered, the evidence for and against these hypotheses, and the conclusion reached. The essay writing instructions are available online.[3] The online environment remained available to each participant during the essay writing, but there was no further communication between participants.

Debriefing included administration of a usability questionnaire, followed by informal discussion with the experimenter of software usability and strategies used during the experimental session. One week after the experimental session, each participant was required to complete an online post-test (described shortly below) before payment was sent.

3.4 Usability Instrument

We used the QUIS 7.0 for usability evaluation of the experimental software. For details, refer to (QUIS, n.d). The QUIS 7.0 has high reliability (Cronbach’s alpha=0.95 and high construct validity (alpha = 0.86) (Chin, Diehl, & Norman, 1988; Harper, Slaughter, & Norman, 1997).

3.5 Post-test design

The post-test was designed as a 20-item 6-choice objective question and answer instrument. It was based on information contained in the ALS-PD articles. These articles included

[3] ibid

distracter information not directly related to the condition, as well as information relevant to the different possible causes of the medical condition. The post-test contained two classes of multiple-choice questions. *Memory* questions could be answered based purely on distracter information that was presented in a single article to a single participant. Since only one participant received the information, half of the memory questions were based on information presented to P1 and half on information presented to P2, enabling us to test for adequate information sharing. *Integrative* questions could only be answered by integrating information that was distributed across articles and participants but needed to be integrated in order to identify the cause of the medical condition. Integrative questions were further divided: *high integration* questions required integration of information presented 5 or more study sessions apart (related to the "inferential span" of Suthers & Hundhausen, 2003; Suthers, Hundhausen et al., 2003); while *low integration* questions were based on information that appeared less than 5 sessions apart. An effective representational tool that is used appropriately should support integration over longer spans than a poor tool or a tool that is not effectively used. Distracter responses to the questions were designed to discriminate different kinds of errors by including the following kinds of responses:

- Logically correct and consistent with materials
- Logically correct, not in the materials, but related to them
- Logically correct, not in the materials, and not related to them
- Logically incorrect, but in the materials
- Logically incorrect, not in the materials, but related to them
- Logically incorrect, not in the materials, and not related to them

The design rationale was that information more intimately tied to the complex explanatory structure that the participants formed while solving the problem would be slower to fade from memory and be easier to recall than information treated as isolated facts (Baddeley, 1997; Craik, 2002). Also, information that had to be integrated across participants would be more likely to be the subject of collaborative discussion, and thus be subject to social associations that would increase long term retention (Fleming & Alexander, 2001). The test design allows us to separate out evidence for information sharing from evidence of integrative elaboration. If the software differed in effectiveness of support for information sharing, we would see this effect in the memory questions as well as the integrative questions. If the software supported information sharing equally well but differed on effectiveness of support for integration and elaboration, we would see this in the integrative questions alone. Since there is a close relationship between collaborative knowledge construction and integration, H1 predicts that a difference will be found on the integration questions in favor of the conditions provided with evidence maps, but not necessarily on the memory questions, as they depend only on information sharing, which can just as well be done in any unstructured but persistent messaging medium. The pragmatic version of H2 predicts that a workspace that integrates discourse and conceptual representations will better support collaborative knowledge construction, in part because of the contextualization of contributions and ease of reference. Again, there is no reason to predict that information sharing is improved over other persistent media, but conceptual integration should be better supported by a single workspace that supports representational integration of all contributions. In contrast, H3 emphasizes separation of discourse and conceptual representations in order to optimize each for their respective purpose. Integration is handled secondarily through a referencing mechanism (the iconic links). Comparison of post-test results for H2 and H3 will determine which strategy is more successful with respect to integration, and hence collaborative knowledge construction.

The post-test has high content validity, as the test is intended to measure sharing and integration of information from the experimental study materials, and all 20 questions were derived from these materials in a manner requiring sharing and for some items integration.

None of the items on the test required that participants have access to or prior knowledge of information outside of the experimental materials, and we verified that they did not have prior knowledge of Guam ALS-PD. Construct validity is addressed by drawing upon empirically validated constructs of memory and integration from the cognitive psychology literature, as discussed above. Reliability is addressed indirectly through validity. Any direct computation of reliability metrics for the post-test is complicated by the fact that the test was administered online one-week after the experimental study in a location of the participant's choosing. We therefore lack the standard conditions needed to compute reliability measures.

3.6 Data collection

Demographic information was collected through a survey and by obtaining Scholastic Aptitude Test (SAT) scores and Grade Point Averages (GPA) from the University (with participants' permission).

Process data was collected through two primary means. First, the Morae™ recording software was used to capture both the computer screen and a webcam sized image of each participant in digital video, along with an audio track. Second, our software was designed to generate complete logs of all the events at each client workstation. These events included message and evidence map object creation, edits, moves, and read events, whether generated by the local or remote participant.

Post-session data included the essay and the usability questionnaire elicited immediately after the experimental session, and the post-test elicited one week later.

4. Results

Our analyses addressed *usability,* based on a usability questionnaire and post-session interviews; *outcomes,* based on several content analyses of the essays and scoring of the post-test; and *session processes,* based on an exploratory examination of the video data to identify recurring issues, an unplanned quantitative analysis conducted to pursue a hypothesis raised by this video analysis, and planned quantitative analyses of elaboration on hypotheses. We describe the results each of these analyses in turn below.

4.1 Usability results

Quantitative analysis of the usability instrument verified that there was no significant difference across groups in participants' satisfaction with the instructions and software demonstration given by the experimenter, providing a check against experimenter bias in these presentations. Analysis of questions pertaining to the software itself yielded a significant difference in satisfaction: Graph received the lowest subjective satisfaction scores and Text the highest. Questions dealing with management of layout of the evidence map representation contributed strongly to this result. Examination of comments confirmed that Graph and Mixed received more negative comments, particularly with respect to screen clutter. Greater prior familiarity with threaded discussion may also have been a factor. In all conditions, participants wanted their contributions to be distinguished from those of their partner. We had elected to make contributions anonymous because we wanted participants to view the workspace as a joint product, but participants felt that visualization of authorship would facilitate their search for information. In the Graph and Mixed conditions, undo was the most requested feature. In the Text conditions, participants frequently asked why they could not interact in real-time with synchronous messaging.

4.2 Outcomes analyses

Analysis of the essays and post-test focused on evidence for information sharing, convergence and quality of hypotheses chosen, and individual memory for facts.

4.2.1 Facts expressed in the essays

A content analysis was conducted on the individually written essays to determine whether the treatment groups differed in the facts expressed in the essays, with particular attention to evidence for sharing of information given to only one participant. We identified propositional facts mentioned in each participant's essay, restricting our attention to information provided by the study materials. Each fact was coded with the name of the article that originally introduced the information. Two analysts independently carried out the analysis and conflicts were resolved by consensus. We then compared conditions on whether participants cited facts that were originally given to the other person (i.e., the information was successfully shared). The results are presented in Table 3. There was no statistically significant difference between the three conditions in the number of facts *mentioned* in individual essays. Also, there was no statistically significant difference in the number of facts *shared* by the two participants. A one-way analysis of variance did not indicate significant differences across the three conditions on participant's preference for facts from their own materials. Participants in the Text condition had more shared facts than Graph and Mixed, possibly because they made more use of copy and paste from messages to the essay, but the difference is not significant. This result provides an important background for subsequent analyses reported below, since other differences between essays cannot be attributed to differences in information sharing.

Table 3: Essay information sharing analysis: A summary

	From P1's Materials	From P2's Materials	Total Facts
Mixed	6.35	6.4	12.75
Graph	6.3	5.95	12.25
Text	6.7	8.25	14.95

	From P1's Materials	From P2's Materials	Total Facts
Mixed P1	7.8	5.7	13.5
Mixed P2	4.9	7.1	12
Graph P1	7.3	4.6	11.9
Graph P2	5.3	7.3	12.6
Text P1	7.5	8.2	15.7
Text P2	5.9	8.3	14.2

	Shared Facts	P1 Facts Shared	P2 Facts Shared
Mixed	3.1	1.1	2
Graph	3.7	2.1	1.6
Text	5.9	2.5	3.4

4.2.2 Solution hypotheses identified in the essays

We examined the hypotheses provided in the essays in response to the instructions: "Write a concluding paragraph in which you identify one or more hypotheses that you believe are *best* supported by the evidence". Two analysts conducted this analysis, obtained similar results, and selected a final analysis by consensus. These hypotheses were compared across treatment conditions to assess differences in *convergence*, as measured by whether each pair's individual essays agree on the cause for the disease (the maximum possible is 10 pairs per condition), and *quality of solution*, as measured by whether individuals identified the most sophisticated and encompassing explanation, namely that the bats were the vector introducing the toxin from cycads into people (the maximum possible is 20 individuals per condition). The results from this analysis are shown in Table 4. From the standpoint of *quality of solution* (under an admittedly simple measure), the difference is not significant under χ^2, although there is a slight trend for more Text participants identifying the relevance of bats. There are clear differences between treatment groups in *convergence of conclusions*, with greater convergence of reasoning in the Graph condition.

Table 4: Conclusions selected in essays

	Convergence (Pair agreement)	Quality (Bat hypothesis)
Text	4/10	5/20
Graph	8/10	2/20
Mixed	2/10	2/20
χ^2	$p \leq 0.025$	$p \leq 1.0$

4.2.3 Post-test results

Recall that the post-test included both memory and integrative questions, and that H1 implies that evidence map users would have an advantage on integrative questions because the evidence map supports greater integration of knowledge, resulting in better memory for precisely the information that was involved this processing. However, no predictions about differences in memory for other information were made.

No significant differences were found in total scores (combining memory and integration questions) across conditions. Comparison of participants' performance on memory for one's own information versus memory for information given to one's partner yielded no statistically significant difference. Therefore, the post-test results provide no evidence for differences between the software conditions in terms of either individual memory or information sharing between participants. This result parallels the lack of differences in number of facts or shared facts expressed in the essays (section 4.2.1).

Table 5: Post-test high integration questions: ANOVA Results

n	60				
High Integration by Condition	n	Mean	SD	SE	
Graph	20	2.200	1.196	0.2675	
Mixed	20	1.200	0.834	0.1864	
Text	20	2.050	1.356	0.3033	
Source of variation	SSq	DF	MSq	F	p
Condition	11.633	2	5.817	4.40	0.0167
Within cells	75.350	57	1.322		
Total	86.983	59			
Contrast	Difference	Bonferroni 95% CI			
Graph v Mixed	1.000	0.103	to 1.897	(significant)	
Graph v Text	0.150	-0.747	to 1.047		
Mixed v Text	-0.850	-1.747	to 0.047		

However, a difference was found on high (but not low) integration questions—those questions requiring integration of information across a span of 5 or more study sessions

(Table 5). This difference was not between the evidence mapping conditions and Text as H1 predicts. Rather, Graph participants performed better than Mixed participants, a result that bears more on the competition between H2 and H3.

4.3 Process analysis of session data

Analyses of the sessions themselves enable us to identify possible explanations for the outcome differences. Although most of these quantitative analyses were planned, we first report on an exploratory qualitative analysis that led to an unplanned quantitative analysis because the latter leads nicely into the planned analyses.

4.3.1 Exploratory examination of session data

Several experimental sessions from each condition were skimmed with the Morae™ video analysis tool, and log data was examined where needed for more precise determination of events. Anticipating the planned analyses, this exploratory analysis focused on the creation, discussion, modification, and referencing of hypotheses.

Our most salient observation concerned the timing and handling of hypotheses. In Graph and Mixed conditions, participants considered the first hypothesis much earlier than in the Text condition. Also, there seemed to be little discussion in the Text condition compared to the other two. For example, many messages were created simply by copying and pasting articles to be shared with the partner. Exploratory analysis suggested that there was little subsequent referencing to hypotheses in the Text condition. In general, substantial discussion of hypotheses in the Text condition took place late in the experimental session. These observations prompted us to conduct a quantitative analysis of the time to create the first hypothesis, in addition to planned analyses of elaboration on hypotheses.

Table 6: Time elapsed before first hypothesis is stated: 1-way between subjects ANOVA results

n	60				
Seconds to the First Hypothesis by Condition	n	Mean	SD	SE	
Graph	20	618.0	568.9	127.20	
Mixed	20	1162.4	1244.3	278.24	
Text	20	2433.8	1807.7	404.22	
Source of variation	SSq	DF	MSq	F	p
Condition	34732575.1	2	17366287.6	10.14	0.0002
Within cells	97656833.5	57	1713277.8		
Total	132389408.6	59			
Contrast	Difference	Bonferroni 95% CI			
Graph v Mixed	-544.5	-1565.5	to 476.6		
Graph v Text	-1815.8	-2836.8	to -794.8	(significant)	
Mixed v Text	-1271.4	-2292.4	to -250.3	(significant)	

4.3.2 Time elapsed to introduction of first hypothesis

A post-hoc test of the time to consider the first hypothesis was motivated by the exploratory analysis. Do the representational tools used differ in how early they encourage participants to state a hypothesis? The analysis measured the time in seconds for each individual participant to introduce the first hypothesis in any medium. A one way-ANOVA conducted on the time in seconds taken to create the first hypothesis yielded significant results ($F(2,57)=10.14$, $p=.0002$). Graph had the earliest creation of the first hypothesis with a mean time of 618 seconds from the start of the first ALS-PD study session. The Mixed condition was ranked next with a mean of 1162 seconds as compared to the Text condition with a mean time of 2433 seconds. The full results are presented in Table 6. The significance of this result is that early introduction of a hypothesis provides more opportunities for elaboration on and evaluation of the hypothesis, but also carries the danger of becoming fixated on a hypothesis. The planned analyses reported next explore how hypotheses were handled during the study sessions.

4.3.3 Consideration of hypotheses during the sessions

H1 predicted that collaborative knowledge construction is more effectively supported by environments that make conceptual relations explicit, because knowledge construction is a process of elaboration and integration that requires awareness of one's own conceptual understanding (i.e., is reflective). An analysis of elaboration and integration was undertaken to test this prediction. For brevity, the remainder of this section uses *elaboration* to include integration. Elaboration was defined to be any action that explicitly considered an already created hypothesis, for example by rewording the hypothesis, discussing the implications of the hypothesis, or providing evidence in support of or against the hypothesis. The analysis encompassed both the contents of linguistic expressions and manipulations of the evidence map, if present. Two coders performed the analysis independently and then the final results were arrived at by consensus. A one-way analysis of variance of the *total elaborations* on hypotheses revealed significant differences between the groups ($F(2, 57)=13.59$, $p<0.0001$). The results are presented in Table 7. There were more elaboration acts in the two treatment conditions that offer an evidence mapping tool: both Graph and Mixed had considerably more elaborative acts than Text.

Further analyses will shape our interpretation of these results. Do some groups consider more hypotheses? A one-way analysis of variance of the number of hypothesis expressed revealed significant differences between the treatment groups ($F(2, 57)=13.59$ $p<0.0001$). The results are presented in Table 8. Participants in Graph expressed significantly more hypotheses than in Text. Therefore the essay convergence result (section 4.2.2) cannot be attributed to Graph pairs simply considering fewer options.

Correlations across groups were also computed during these analyses. As we hypothesized, there was a negative correlation ($r=-0.38$, $p=.003$) between the time taken to create the first hypothesis (M=1404.07 seconds, SD=1497.96, N=60) and total elaborations (M=11.33, SD=10.78, N=60) in an experimental session. A similar analysis revealed a positive correlation ($r=0.59$, $p<.0001$) between total hypotheses (M=4.37, SD=2.60, N=60) and total elaborations (M=11.33, SD=10.78, N=60) in the experimental session.

Is there more elaboration in Graph simply because there are more hypotheses, or does each hypothesis receive greater consideration? A one-way analysis of variance of the average number of elaborations per hypothesis was significant (Table 9). The differences are between both Graph and Mixed versus Text: the presence of an evidence mapping tool results in more elaboration on each idea considered.

Table 7: Total elaborations: ANOVA results

n	60				
Total Elaborations by Condition	n	Mean	SD	SE	
Graph	20	17.900	13.742	3.0727	
Mixed	20	12.850	7.051	1.5766	
Text	20	3.250	2.447	0.5471	
Source of variation	SSq	DF	MSq	F	p
Condition	2215.233	2	1107.617	13.59	<0.0001
Within cells	4646.100	57	81.511		
Total	6861.333	59			
Contrast	Difference	95% CI			
Graph v Mixed	5.050	-1.820	to 11.920		
Graph v Text	14.650	7.780	to 21.520	(significant)	
Mixed v Text	9.600	2.730	to 16.470	(significant)	

Table 8: Total hypotheses: ANOVA results

n	60				
Total Hypothesis by Condition	n	Mean	SD	SE	
Graph	20	5.7	3.1	0.69	
Mixed	20	4.2	2.4	0.54	
Text	20	3.3	1.7	0.37	
Source of variation	SSq	DF	MSq	F	p
Condition	56.6	2	28.3	4.73	0.0126
Within cells	341.3	57	6.0		
Total	397.9	59			
Contrast	Difference	Tukey 95% CI			
Graph v Mixed	1.5	-0.4	to 3.4		
Graph v Text	2.4	0.5	to 4.2	(significant)	
Mixed v Text	0.9	-1.0	to 2.7		

Table 9: Elaborations per hypothesis: ANOVA results

n	60				
Elaborations per Hypothesis by Condition	n	Mean	SD	SE	
Graph	20	3.785	3.634	0.8127	
Mixed	20	3.781	2.981	0.6666	
Text	20	0.995	0.762	0.1703	
Source of variation	SSq	DF	MSq	F	p
Condition	103.653	2	51.827	6.86	0.0021
Within cells	430.849	57	7.559		
Total	534.502	59			
Contrast	Difference	Bonferroni 95% CI			
Graph v Mixed	0.004	-2.141	to 2.148		
Graph v Text	2.790	0.645	to 4.935	(significant)	
Mixed v Text	2.786	0.642	to 4.931	(significant)	

The Graph and Mixed software provide both evidence mapping and text-oriented tools. Is the increased elaborative activity attributable to the evidence mapping affordances? Another analysis counted elaborations in the textual representations alone (the threaded discussions and the note objects in the evidence map). The test for significance did not meet the criteria of $p<.05$ (Table 10), although there was a trend for more elaboration in Text, which is not surprising since they had no where else to do it. Therefore, the increased elaboration observed in Graph and Mixed takes place in and is attributable to the evidence mapping representations, even though text-only discourse representations are available in all groups.

Table 10: Discussion elaborations only: ANOVA results

n	60				
Discussion Elaborations by Condition	n	Mean	SD	SE	
Graph	20	1.800	1.852	0.4142	
Mixed	20	1.950	1.905	0.4260	
Text	20	3.250	2.447	0.5471	
Source of variation	SSq	DF	MSq	F	p
Condition	25.433	2	12.717	2.92	0.0618
Within cells	247.900	57	4.349		
Total	273.333	59			
Contrast	Difference	Tukey 95% CI			
Graph v Mixed	-0.150	-1.737	to 1.437		
Graph v Text	-1.450	-3.037	to 0.137		
Mixed v Text	-1.300	-2.887	to 0.287		

5. Discussion

We now review the hypotheses stated at the outset in terms of the results, beginning with the primary hypothesis:

> *H1: Collaborative knowledge construction is more effectively supported by environments that make conceptual objects and relations explicit.*

Two lines of evidence support this hypothesis, based on process and outcome data. The process data shows clearly that there was more elaboration on hypotheses in both of the environments that made conceptual objects and relations explicit (Graph and Mixed) as compared to the environment that did not (Text). Hypotheses were stated earlier in the experimental session (section 4.3.2) and there was more elaboration on the hypotheses individually as well as collectively (section 4.3.3). Furthermore, Graph users considered more hypotheses (section 4.3.3). These results are consistent with the representational guidance effect demonstrated for face-to-face interaction in a laboratory setting by Suthers and Hundhausen (2003) and in a classroom setting by (Toth et al., 2002). See also (Veerman, 2003) for a related study in a synchronous online setting. In summary, process measures suggest that more knowledge construction takes place when interaction is supported by conceptual representations.

Although the process analyses did not specifically consider group processes, the outcome data suggests that there are consequences at the group level. The analysis of solution hypotheses identified in the essays (section 4.2.2) showed that participants in Graph were more likely to converge, expressing the same conclusions in their essays. This convergence cannot be attributed to a paucity of alternatives: the process data shows that Graph users considered *more* hypotheses than the others, which makes their convergence even more notable. The convergence is probably not due to more effective information sharing per se, since there were no differences on facts mentioned in the essay (section 4.2.1) or on memory for information given to one's partner (section 0). (Information sharing will be discussed further below.) A plausible explanation is that the shared and visually oriented evidence mapping workspace (which was available during the essay writing) enables participants to both see the same "big picture" from which they draw the same conclusions while writing the essays—a "group mirror" (Dillenbourg, 2005). This explanation admits the possibility that convergence took place only during essay writing rather than the sessions. Yet, the same evidence mapping workspaces were also shared during the session, so the same argument can be made for the role of the visual workspace in coordinating collaborative activity. Given the process data just reviewed, it is plausible that collaborative consideration of hypotheses *during* the study sessions had an effect on convergence of the participants' conclusions. A study of face-to-face collaboration (Suthers & Hundhausen, 2003) similarly found that the work done with an evidence map representation during study sessions had greater bearing on essay contents than the work done with a matrix or a text representation. The similarity of results is interesting in light of the differences between these studies: in addition to the media difference, Suthers and Hundhausen's participants wrote *collaborative* essays *from memory*.

On the other hand, if "more effectively" means "producing better outcomes," then the lack of differences on quality of solution (section 4.2.2) and overall post-test performance (section 0) may be counted as evidence against H1. The slightly greater identification of the bats hypothesis by Text participants (Table 4) might reflect the tendency of the Text participants to simply cut and paste entire articles into their text messages and leave discussion for the end. Evidence for the bat hypothesis was provided near the end of the experimental session, so the final set of messages available in the sequential representation would be more likely to mention this hypothesis, and therefore would be more likely to be pasted into the essay (a recency bias; Hewitt, 2003).

The failure of the Mixed condition in some analyses to display the advantages claimed by H1 may also be considered as evidence against H1. Participants in the Mixed condition may have converged the least because the dual workspaces of Mixed provide more variation in strategies for using the workspaces, increasing the possibility that members of a pair will look at different material. An explanation as to why Mixed performed the worst on high integration questions is offered in the discussion of H3, below.

In summary, the hypothesis elaboration and convergence results provide support for H1, although solution quality and post-test results are equivocal and the results depend on the software design chosen. This possibility was anticipated: the realization that implementation choices may influence the viability of H1 motivated H2 and H3 and our decision to test both Graph and Mixed designs. These two hypotheses will be discussed together:

> *H2: Collaborative knowledge construction is more effectively supported if discourse and conceptual representations are tightly integrated.*

> *H3: Collaborative knowledge construction is more effectively supported if the distinction between discourse and conceptual is reflected in the representations provided.*

Direct evidence on which H2 (Graph) and H3 (Mixed) may be compared is limited. The only significant difference we found between these two treatment conditions was that Graph users remember more integrative relationships than Mixed (section 0). Our prior work (Suthers, Girardeau, & Hundhausen, 2003; Suthers & Hundhausen, 2003) showed that evidence mapping encourages references to and integration with information introduced earlier in time, an important advantage in light of the propensity for users of threaded discussion to reply primarily to recently posted messages and hence not resolve important issues (Hewitt, 2003). Why wasn't this advantage observed in both evidence mapping conditions of the present study? The additional complexity of using two representations (the threaded discussion and the evidence map) may have been a factor in Mixed (Ainsworth, Bibby, & Wood, 1998). Although the visibility and conceptual organization of information in

Graph's evidence map appears to have supported integration as predicted by H1, the distribution of information across two media in Mixed may have posed a barrier to integration of that information, obscuring the advantage of Mixed's evidence map.

There is indirect evidence bearing on the choice between H2 and H3. All other statistical analyses in which there was a significant advantage for one of the conditions over the others included an advantage of Graph over Text, with the exception of the subjective assessment of usability. In contrast, Mixed was sometimes advantageous to Text, sometimes not, but never was advantageous to Graph, and sometimes yielded the worst results. Therefore, in reference to the alternative of threaded discussions, support for H2 is stronger than for H3.

This result is at odds with an earlier prediction (Suthers, 2001) that a logically linked relationship between distinct discourse and knowledge representations would be preferable to an embedded approach. The present (embedded) Graph environment does provide some of the optimizations that Suthers (2001) claimed were associated with linked representations: the note objects function as mini-discussions, preserving local chronological structure, while enabling referencing through both literal links and spatial proximity in the evidence map. No global representation of the chronology of discussion is provided, but nor did the task posed to participants in this study particularly require a chronological organization.

6. Summary and Conclusions

Many tools for online collaborative learning are text based, typically providing representational support only for discourse structure in the form of reply relations (threading) of contributions. Along with others (Turoff et al., 1999), we have argued that tools for online learning should provide representational support for conceptual structure in order to address issues of coherence and convergence and more effectively support collaborative knowledge construction (Suthers, 2001). The study described in this paper set out to investigate the claimed merits of conceptually oriented representations and of two approaches to the relationship between conceptual and discourse representations: embedded or linked. This study was undertaken in an asynchronous setting, using a protocol for practical experimental study of asynchronous collaboration in the laboratory. It is crucial to understand how to support collaborative knowledge construction in asynchronous settings due to the prevalence of asynchronous approaches to online learning. Therefore we hope that other researchers working in the experimental paradigm will take advantage of the quasi-asynchronous protocol.

The results included process and outcome differences. A process analysis focused on the creation and discussion of hypotheses. A representational effect was identified: users of a knowledge representation tool that includes primitives for hypotheses are more likely to state hypotheses early in their experimental sessions, elaborate on these hypotheses and integrate them with data than users of the threaded discussion tool. In the threaded discussion, participants tended to simply record the literal text of the information articles, and not discuss hypotheses until later in the experimental session. Examination of the final conclusions stated in the essays shows that pairs using the evidence map with embedded annotations were more likely to converge on the same hypothesis, even though more hypotheses had been considered, and results from a post-test conducted a week later also suggested that embedded conceptual representations improve collaborative integration of information.

There is indirect evidence that the operative mechanism was not differences in information sharing. This evidence is indirect because it is based on outcome data: an analysis of facts expressed in the essays and an analysis of memory for facts that a participant would have to obtain from his or her partner. An analysis that traced out information sharing during the session would provide more direct evidence. Such an analysis is underway as this article goes to press, and will be reported in future publications. The analysis is of interest because information sharing is at the center of the knowledge-communication paradigm (Wenger, 1987). Studies of information sharing remain vital and relevant (Bromme, Hesse, & Spada, 2005), yet the present study suggests that the view that "the gist of cooperative learning [is] going from unshared to shared information" (Pfister, 2005) is missing something important. Something different may be happening *interactionally* in Graph beyond information sharing.

In general, it is time to intensify our study of what happens interactionally in groups: the "distributed cognition" (Hutchins, 1995), "intersubjective meaning-making" (Suthers, 2006b) or "group cognition" (Stahl, 2006) that emerges from interacting individual cognitions. We mentioned at the outset of this paper that, in addition to our hypothesis-testing objectives, this study was designed to provide rich data that would enable analysis of media appropriation in collaborative knowledge construction. Towards this end, we are presently developing a methodology for analysis of intersubjective meaning-making in asynchronous contexts and applying it to data from the present study (Suthers, 2006a; Suthers, Dwyer, Vatrapu, & Medina, 2007).

The primary finding of this study—that collaborative knowledge construction is fostered by conceptual representations—not only adds to the growing literature on representational guidance for collaborative learning, but also has practical implications. Should threaded discussion tools be replaced with knowledge mapping tools in online learning? Although that is the direction in which this study points, it would be a brash conclusion to draw from this study alone, as it is limited in many ways. We studied dyads interacting over a relatively short period of two hours. Dozens of students interacting over the course of a semester (even if divided into smaller groups as is generally recommended in ALN implementations) would generate much more complex artifacts, greatly exacerbating the problem exposed by the usability questionnaire and immediately apparent from some of the complex artifacts our participants produced. Any workspace has a limited useful life before it becomes important to "rise above" the clutter and start fresh, as the CSILE project has shown (Scardamalia, 2004). The subject matter, task structure, and nature of the representations used could also affect results. Clearly, there are ample opportunities for further research. However, we believe that in conjunction with previous work the present results merit extending the research program beyond the laboratory by undertaking action research in which richer interactive representations are studied in settings of educational practice.

7. Acknowledgments

David Burger and Niels Pinkwart have contributed to the design of this study and the implementation of the software on which it is based. Arlene Weiner introduced the first author to the "science challenge problems," which were reworked substantially for this study. This work was supported by the National Science Foundation under CAREER award 0093505.

8. References

Ainsworth, S. E., Bibby, P. A., & Wood, D. J. (1998). Analyzing the costs and benefits of multi-representational learning environments. In M. W. van Someren, P. Reimann, H. P. A. Boshuizen & T. de Jong (Eds.), *Learning with Multiple Representations* (pp. 120-134). Amsterdam: Elsevier Science, Ltd.

Allen, I. E., & Seaman, J. (2005). Growing by Degrees: Online Education in the United States, 2005. *The Sloan Consortium*, http://www.sloan-c.org/resources/growing_by_degrees.pdf.

Baddeley, A. (1997). Human Memory: Theory and Practice. In (pp. 115-123): Psychology Press,.
Baker, M. J., & Lund, K. (1996). Flexibly Structuring the Interaction in a CSCL environment. In P. Brna, A. Paiva & J. Self (Eds.), *Proceedings of the EuroAIED Conference* (pp. 401-407). Edições Colibri, Lisbon.
Bell, P., Davis, E. A., & Linn, M. C. (1995). The Knowledge Integration Environment: Theory and design, *Proceedings of the Computer Supported Collaborative Learning Conference '95* (pp. 14-24). Toronto: Lawrence Earlbaum Associates.
Brna, P., Cox, R., & Good, J. (2001). Learning to think and communicate with diagrams: 14 questions to consider. *Artificial Intelligence Review, 15*(1-2), 115-134.
Bromme, R., Hesse, F. W., & Spada, H. (Eds.). (2005). *Barriers and Biases in Computer-Mediated Knowledge Communication—and How They May Be Overcome*. Dordrecht: Kluwer.
Brown, A. L., & Campione, J. C. (1994). Guided discovery in a community of learners. In K. McGilly (Ed.), *Classroom Lessons: Integrating Cognitive Theory and Practice* (pp. 229-270). Cambridge:: MIT Press.
Chi, M. T. H., Bassok, M., Lewis, M., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science, 13*, 145-182.
Chin, J. P., Diehl, V. A., & Norman, K. L. (1988). Development of an instrument measuring user satisfaction of the human-computer interface, *SIGCHI '88* (pp. 213-218). New York: ACM.
Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. In L. B. Resnick, J. M. Levine & S. D. Teasley (Eds.), *Perspectives on Socially Shared Cognition* (pp. 127-149): American Psychological Association.
Craik, F. I. M. (2002). Levels of processing: Past, present ... and future? *Memory, 10*(5/6), 305-318.
Dillenbourg, P. (2005). Designing biases that augment socio-cognitive interactions. In R. Bromme, F. W. Hesse & H. Spada (Eds.), *Barriers and Biases in Computer-Mediated Knowledge Communication-and How They May Be Overcome*. Dordrecht: Kluwer.
Dwyer, N., & Suthers, D. D. (2005). A Study of the Foundations of Artifact-Mediated Collaboration. In T. Koschmann, D. S. & & T. W. Chan (Eds.), *Computer Supported Collaborative Learning 2005: The Next 10 Years!* (pp. 135-144). Mahwah, NJ: Lawrence Erlbaum Associates.
Dwyer, N., & Suthers, D. D. (in press). Consistent Practices in Artifact-Mediated Collaboration. *International Journal of Computer Supported Collaborative Learning, 1*(4).
Fleming, V. M., & Alexander, J. M. (2001). The benefits of peer collaboration: A replication with a delayed posttest. *Contemporary Educational Psychology, 26*, 588-601.
Frank, M. (2006). How to teach using today's technology: Matching the teaching strategy to the e-learning approach. In L. T. W. Hin & R. Subramaniam (Eds.), *Handbook of Research on Literacy in Technology at the K-12 Level* (pp. 372-393). Hershey, PA: Idea Group Publishing.
Garfinkel, H. (1967). *Studies in Ethnomethodology*. Englewood Cliffs, New Jersey: Prentice-Hall.
Grosz, B. J., & Sidner, C. L. (1986). Attention, Intentions, and the Structure of Discourse. *Computational Linguistics, 12*(3), 175-204.
Guzdial, M., & Hmelo, C. (1997). Integrating and guiding collaboration: Lessons learned in computer-supported collaborative learning research at Georgia Tech, *Proceedings of Computer Supported Collaborative Learning '97* (pp. 91-100). Toronto, Ontario.

Harper, B., Slaughter, L., & Norman, K. (1997). Questionnaire administration via the WWW: A validation and reliability study for a user satisfaction questionnaire. In S. Lobodzinski & I. Tomek (Eds.), *Proceedings of WebNet 97 - World Conference on the WWW, Internet & Intranet*. Chesapeake, VA: AACE.

Hawkes, M., & Romiszowski, A. (2001). Examining the reflective outcomes of asynchronous computer-mediated communication on inservice teacher development. *Journal of Technology and Teacher Education, 9*(2), 285-308.

Heckman, R., & Annabi, H. (2003). *A Content Analytic Comparison of FTF and ALN Case-Study Discussions.* Paper presented at the Proceedings of the 36th Hawai`i International Conference on the System Sciences (HICSS-36), January 6-9, 2003, Waikoloa, Hawai`i (CD-ROM).

Herring, S. C. (1999). Interactional Coherence in CMC. *Journal of Computer Mediated Communication, 4*(4).

Hewitt, J. (2001). Beyond Threaded Discourse. *International Journal of Educational Telecommunications, 7*(3), 207-221.

Hewitt, J. (2003). How habitual online practices affect the development of asynchronous discussion threads. *Journal of Educational Computing Research, 28*(1), 31-45.

Hill, J. R., Wiley, D., Nelson, L. M., & Han, S. (2003). Exploring research on Internet-based learning: From infrastructure to interactions. In D. H. Jonassen (Ed.), *Handbook of Research for Educational Communications and Technology* (2nd ed., pp. 433-460). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Hiltz, S. R., Coppola, N., Rotter, N., & Turoff, M. (2000). Measuring the Importance of Collaborative Learning for the Effectiveness of ALN: A Multi-Measure, Multi-Method Approach. *Journal of Asynchronous Learning Networks, 4*(2), http://www.sloan-c.org/publications/jaln/v4n2/v4n2_hiltz.asp.

Hoadley, C., & Enyedy, N. (1999). Between information and communication: Middle spaces in computer media for learning. In C. Hoadley & J. Roschelle (Eds.), *Computer Support for Collaborative Learning (CSCL)* (pp. 242-251). Palo Alto, CA: Stanford University.

Hutchins, E. (1995). *Cognition in the Wild*. Cambridge, Massachusets: The MIT Press.

Kitade, K. (2000). L2 learners' discourse and SLA theories in CMC: Collaborative interaction in Internet chat. *Computer Assisted Language Learning, 13*(2), 143-166.

Koedinger, K. (1991). On the design of novel notations and actions to facilitate thinking and learning, *Proceedings of International Conference on the Learning Sciences* (pp. 266-273). Charlottesville, VA: Association for the Advancement of Computing in Education.

Kotovsky, K., & Simon, H. A. (1990). What makes some problems really hard: Explorations in the problem space of difficulty. *Cognitive Psychology,, 22*, 143-183.

Larkin, J. H., & Simon, H. A. (1987). Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science, 11*(1), 65-99.

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge: Cambridge University Press.

Lieberman, A. (2004). What You Should Know About Guam, Plants, Flying Foxes and Parkinson Disease: National Parkinson Foundation.

Mayadas, F. (1997). Asynchronous Learning Networks: A Sloan Foundation Perspective. *Journal of Asynchronous Learning Networks, 1*, http://www.aln.org/alnweb/journal/jaln_issue1.htm#mayadas.

Mühlpfordt, M., & Wessner, M. (2005). Explicit Referencing In Chat Supports Collaborative Learning. In T. Koschmann, D. Suthers & T. W. Chan (Eds.), *Computer Supported*

Collaborative Learning: The Next 10 Years! (pp. 460-469). Mahwah, NJ: Lawrence Erlbaum Associates.

Novak, J. (1990). Concept mapping: A useful tool for science education. *Journal of Research in Science Teaching, 27*(10), 937-949.

Novick, L. R., & Hmelo, C. E. (1994). Transferring symbolic representations across nonisomorphic problems. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*(6), 1296-1321.

Okada, T., & Simon, H. A. (1997). Collaborative discovery in a scientific domain. *Cognitive Science, 21*(2), 109-146.

Perkins, D. N., Crismond, D., Simmons, R., & Unger, C. (1997). Inside understanding. In D. N. Perkins, J. L. Schwartz, M. M. West & M. S. Wiske (Eds.), *Software Goes to School: Teaching for Understanding with New Technologies* (pp. 70-87). New York: Oxford University Press.

Pfister, H.-R. (2005). How to support synchronous net-based learning discourses: Principles and perspectives. In R. Bromme, F. Hesse & H. Spada (Eds.), *Barriers and Biases in Computer-Mediated Knowledge Communication--And How They May be Overcome*. Dordrecht: Kluwer.

QUIS. (n.d). *Questionnaire for User Interaction Satisfaction*. Retrieved May 22, 2006, from http://www.lap.umd.edu/QUIS/index.html

Reyes, P., & Tchounikine, P. (2003). Supporting Emergence of Threaded Learning Conversations Through Augmenting Interactional and Sequential Coherence. In B. Wasson, S. Ludvigsen & U. Hoppe (Eds.), *Designing for Change in Networked Learning Environments-Proceedings of Conference CSCL 2003* (pp. 83-92). Dordrecht: Kluwer Academic Publishers.

Roschelle, J. (1996). Designing for cognitive communication: Epistemic fidelity or mediating collaborating inquiry. In D. L. Day & D. K. Kovacs (Eds.), *Computers, Communication & Mental Models* (pp. 13-25). London: Taylor & Francis.

Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language, 50*(4), 696-735.

Scardamalia, M. (2004). CSILE/Knowledge Forum®. In *Education and technology: An encyclopedia.* (pp. 183-193). Santa Barbara: ABC-CLIO.

Scardamalia, M., & Bereiter, C. (1991). Higher levels of agency for children in knowledge building: A challenge for the design of new knowledge media. *The Journal of the Learning Sciences, 1*(1), 37-68.

Scardamalia, M., & Bereiter, C. (1994). Computer support for knowledge building communities. *Journal of the Learning Sciences, 3*(3), 265-283.

Scardamalia, M., Bereiter, C., Brett, C., Burtis, P. J., Calhoun, C., & Smith Lea, N. (1992). Educational applications of a networked communal database. *Interactive Learning Environments, 2*(1), 45-71.

Snir, J., Smith, C., & Grosslight, L. (1995). Conceptually enhanced simulations: A computer tool for science teaching. In D. N. Perkins, J. L. Schwartz, M. M. West & M. S. Wiske (Eds.), *Software Goes to School: Teaching for Understanding with New Technologies* (pp. 106-129). New York: Oxford University Press.

Stahl, G. (2006). *Collaborating with Technology: Mediation of Group Cognition*. Cambridge, MA: MIT Press.

Stasser, G., & Stewart, D. (1992). Discovery of hidden profiles by decision-making groups: Solving a problem versus making a judgment. *Journal of Personality and Social Psychology Quarterly, 57*, 67-78.

Sun, C., Jia, X., Zhang, Y., & Yang, Y. (1997). A generic operation transformation scheme for consistency maintenance in real-time cooperative editing systems, *Proceedings of ACM Conference on Supporting Group Work* (pp. 425-434).

Suthers, D. D. (2001). Collaborative Representations: Supporting Face to Face and Online Knowledge-building Discourse. In *Proceedings of the 34th Hawai`i International Conference on the System Sciences (HICSS-34), January 3-6, 2001, Maui, Hawai`i (CD-ROM)*: Institute of Electrical and Electronics Engineers, Inc. (IEEE).

Suthers, D. D. (2006a). A qualitative analysis of collaborative knowledge construction through shared representations. To appear in *Research and Practice in Technology Enhanced Learning 1*(2).

Suthers, D. D. (2006b). Technology affordances for intersubjective meaning-making: A research agenda for CSCL. *International Journal of Computers Supported Collaborative Learning, 1*(3), 315-337.

Suthers, D. 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. F. a. P. J. Feltovich (Ed.), *Smart Machines in Education: the Coming Revolution in Educational Technology* (pp. 7-35). Cambridge, Massachusetts: AAAI Press, The MIT Press.

Suthers, D. D., Dwyer, N., Vatrapu, R., & Medina, R. (2007). An abstract transcript notation for analyzing interactional construction of meaning in online learning. In *Proceedings of the 40th Hawai`i International Conference on the System Sciences (HICSS-40), January 3-6, 2007, Waikoloa, Hawai`i (CD-ROM)*: Institute of Electrical and Electronics Engineers, Inc. (IEEE).

Suthers, D. 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.), *Proceedings of the International Conference on Computer Support for Collaborative Learning 2003* (pp. 173-182): Dordrecht: Kluwer Academic Publishers.

Suthers, D. D., & Hundhausen, C. (2003). An Experimental Study of the Effects of Representational Guidance on Collaborative Learning. *Journal of the Learning Sciences, 12*(2), 183-219.

Suthers, D. D., Hundhausen, C., & Girardeau, L. (2003). Comparing the roles of representations in face-to-face and online computer supported collaborative learning. *Computers and Education, 41*, 335-351.

Tabak, I. (2004). Synergy: A complement to emerging patterns of distributed scaffolding. *Journal of the Learning Sciences, 13*(3), 205-335.

Toth, E. E., Suthers, D. D., & Lesgold, A. M. (2002). "Mapping to Know": The Effects of Representational Guidance and Reflective Assessment on Scientific Inquiry: Wiley Periodicals, Inc.

Turoff, M., Hiltz, S. R., Bieber, M., Fjermestad, J., & Rana, A. (1999). Collaborative discourse structures in computer mediated group communications. *Journal of Computer Mediated Communication, 4*(4), http://jcmc.huji.ac.il/.

van der Pol, J., Admiraal, W., & Simons, P. R. J. (2006). The affordance of anchored discussion for the collaborative processing of academic texts. *Comptuer-Supported Collaborative Learning, 1*(3), 339-357.

Veerman, A. (2003). Constructive discussions through electronic dialogue. In J. Andriessen, M. Baker & D. D. Suthers (Eds.), *Arguing to Learn: Confronting Cognitions in Computer-Supported Collaborative Learning Environments* (pp. 117-143). Dordrecht: Kluwer.

Von Glasersfeld, E. (1995). A constructivist approach to teaching In L. Steffe & J. Gale (Eds.), *Constructivism in Education* (pp. 3-16). New Jersey: Lawrence Erlbaum Associates, Inc.

Vygotsky, L. S. (1978). *Mind in society*. Cambridge, MA: Harvard University Press.

Webb, N., & Palincsar, A. (1996). Group processes in the classroom. In D. Berlmer & R. Calfee (Eds.), *Handbook of Educational Psychology*. New York: Simon & Schuster Macmillian.

Wells, G. (1999). *Dialogic inquiry: Toward a sociocultural practice and theory of education*. New York: Cambridge University Press.

Wenger, E. (1987). *Artificial Intelligence and Tutoring Systems: Computational and Cognitive Approaches to the Communication of Knowledge*. Los Altos: Morgan Kaufmann.

Zhang, J. (1997). The nature of external representations in problem solving. *Cognitive Science, 21*(2), 179-217.