Conceptual Representations Enhance Knowledge Construction in Asynchronous Collaboration

Daniel Suthers, Ravi Vatrapu, Richard Medina, Sam Joseph, Nathan Dwyer
Laboratory for Interactive Learning Technologies
Dept. of Information and Computer Sciences, University of Hawai`i at Manoa
collaborative-representations@hawaii.edu

Abstract: An experimental study of asynchronously communicating dyads tested the claim that conceptual representations could more effectively support collaborative knowledge construction in online learning than threaded discussions. Results showed that users of conceptual representations created more hypotheses earlier in the experimental sessions and elaborated on hypotheses more than users of threaded discussions. Participants using conceptual representations were more likely to converge on the same conclusion and scored higher on post-test questions that required integration of information distributed across dyads in a hidden profile design. However, the essay contents and post-test offered no evidence for differences in information sharing in itself. 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.

Introduction

Prior work has established the potential value of representational guidance for social processes of learning (Dillenbourg, 2005; Suthers & Hundhausen, 2003). Research on representations that are constructed by learners during collaboration (Suthers & Hundhausen, 2003) and representations used as a discussion medium (Baker & Lund, 1996; Guzdial & Hmelo, 1997) has shown that the choice of representation can change the focus of learning discourse. Research on computer-mediated communication (CMC) has identified problems as well as opportunities related to typical representations through which people communicate online (e.g, threaded discussion and chat). Although discussion forums may support more reflective contributions (Hawkes & Romiszowski, 2001), online interaction can also suffer from *incoherence* due to the violation of adjacency conventions for topic maintenance (Herring, 1999) and the coarse granularity of referencing (Reyes & Tchounikine, 2003). Furthermore, there can be a *lack of convergence* due to the intrinsically divergent representations used in threaded discussion (Hewitt, 2001) and a bias towards addressing recently posted messages (Hewitt, 2003). The shared knowledge being constructed 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 discussion (Suthers, 2001; Turoff, Hiltz, Bieber, Fjermestad, & Rana, 1999).

Suthers (2001) argued that if the conceptual development of the conversation can be made explicit and each contribution to the discussion can be referenced to a component of this conceptual representation, coherence may improve because the conceptual relevance of each contribution is clear (see also van der Pol, Admiraal, & Simons, 2006), and convergence may improve because multiple contributions referencing a given topic are collected together. The present study constitutes an experimental test of these ideas, conducted in an asynchronous setting to inform this increasingly prevalent form of online learning (Mayadas, 1997). Participants were enabled to construct explicit representations of the topics and conclusions of their discussion as they interacted. Two forms of conceptually-enhanced support were compared to each other and to a threaded discussion control condition. Below, we first specify our research hypotheses and explain how these are reflected in the software designs that define the experimental treatments. The remaining sections follow the traditional presentation sequence.

Hypotheses

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 is accomplished when these processes take place in joint as well as individual acts of meaning-making (Scardamalia & Bereiter, 1994; Stahl, 2006). Our primary hypothesis (H1) claims that *collaborative knowledge construction is more effectively supported by environments that make conceptual objects and relations explicit.* (A visual representation of reply structure, as in CSILE (Scardamalia, 2004), does not meet this definition.) 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), can guide students' argumentation to include disconfirming as well as confirming evidence (Toth, Suthers, & Lesgold, 2002; Veerman, 2003), and can function as a "convergence artifact" that expresses the group's emerging consensus (Hewitt, 2001; Suthers, 2001). This primary hypothesis does not specify the relationship between knowledge representations and the conversation that accompanies the creation of those representations. Our secondary hypotheses are alternative elaborations of H1, arguing for either maintaining the distinction between discussion and knowledge representations or combining the two.

One could argue that discussion 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). A usability argument can also be made: it may be easier to manage a single workspace than interactions distributed across multiple tools. This reasoning leads to the second hypothesis (H2): *Collaborative knowledge construction is more effectively supported if conversational and conceptual representations are tightly integrated.*

The third hypothesis is motivated by the observation that conversational structures and conceptual structures are different: conversation 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 conversation and conceptualization in a given domain of discourse. Explicit referencing can be used to make the connection between the two representations (Mühlpfordt & Wessner, 2005; Suthers, 2001). This reasoning leads us to the third hypothesis H3, which is in opposition to the second: *Collaborative knowledge construction is more effectively supported if the distinction between discussion and conceptual models is reflected in the representations provided.*

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 task are displayed. All three environments have a shared workspace or "information organizer" on the right hand side (and in one case the lower left) in which participants can share 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 an asynchronous protocol to be discussed.

The shared workspace in the *Text* condition is a conventional threaded discussion tool (Figure 1). This is 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).
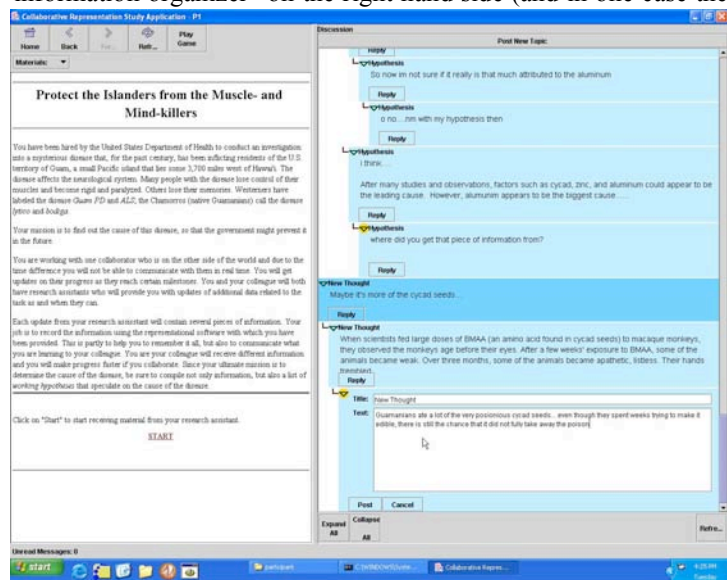

Figure 1. *Text* environment (threaded discussion)

The shared workspace for the *Graph* condition includes tools for constructing conceptual objects under a 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* and *inconsistency* 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 (lower right of 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, as suggested by H2. In this paper, we use *evidence map* to refer to the specific representational tool used in the experiment, and *knowledge map* to refer to the category of conceptually explicit representations.
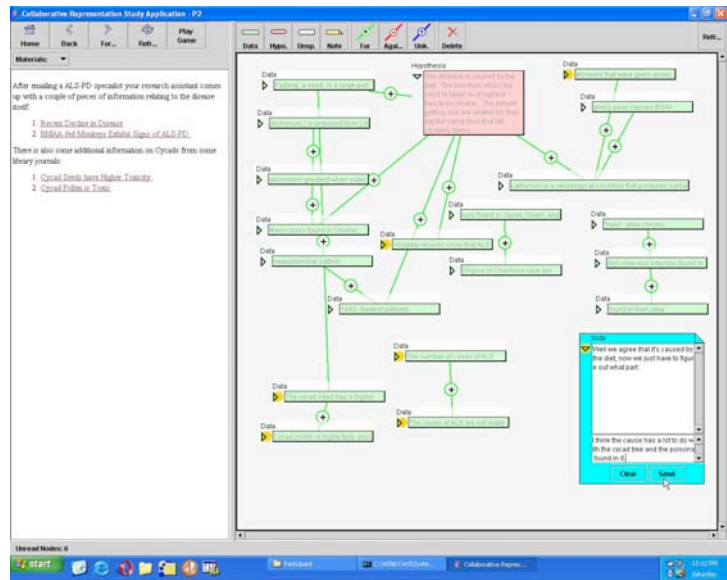
The shared workspace of the *Mixed* condition includes both a threaded discussion tool and an evidence-mapping tool for representing conceptual structure in the same manner as the Graph condition, except that there are no embedded notes in the Mixed version of the evidence map. Instead, one can embed references to evidence map objects in the threaded discussion messages 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. This environment is motivated by H3, which claims that separate representations are needed to optimize discussion and conceptual organization.
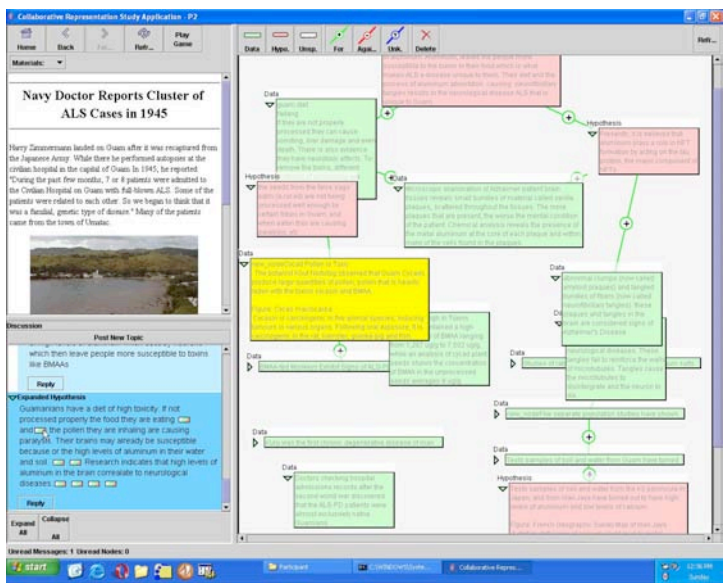

Figure 2. *Graph* environment (knowledge map)

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

Methods

This section provides a brief summary of the experimental method. Further details are forthcoming in (Suthers, Vatrapu, Medina, Joseph, & Dwyer, in press).

Design

H1 predicts that the presence of a conceptual representation will be beneficial. 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. Therefore, in order to determine whether *some* implementation of a conceptual representation is better than threaded discussion alone, we 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. 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.

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 experiment. We recruited participants in pairs of acquaintances so as to eliminate the social awkwardness of interaction between persons who do not know each other. 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: 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 $(F(2, 54) = 0.18, p= 0.8361)$ and grade point average $(F(2, 54) = 1.20, p= 0.3105)$. We also verified through a pre-experiment questionnaire that none of the participants had prior experience with the experimental problem.

Materials

The experiment presented participants with "science challenge" problems, consisting of issues in science and public health. The main 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 60 years, in part because it shares symptoms with Alzheimer's and Parkinson's diseases (Lieberman, 2004). Over the years several hypotheses have been proposed and evaluated with evidence of varying types and quality. 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.

The source materials were divided into twelve (12) sets of materials, each set consisting of a brief introduction and links to four articles. (A complete list of source materials for the ALS-PD problem is available online at http://lilt.ics.hawaii.edu/lilt/papers/2007/Suthers-et-al-CSCL-2007/.) In a given experimental session, each of two participants (designated P1 and P2) was assigned half of these of materials, presented in six "study sessions." Each article typically consisted of one to two brief paragraphs and an image or two. An example article is shown in the left hand side of Figure 3 (shown previously). Each article was designed to provide one key item of information relevant to a hypothesis. The remaining information in a given article elaborated on this item or provided tangentially related "distracter" information. We designed the articles to provide evidence both for (+) and against (-) five major hypotheses (the codes are used in Table 1): (A) aluminum 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).

Table 1: Distribution and sequencing of information articles across participants and study 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+

We used a "hidden profile" (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. The sequencing of articles was designed to require integration over time, motivating use of the information organizing workspace to make relevant information available in later study sessions. Table 1 shows the complete distribution of materials across the participants. For example, one participant (P1) received evidence for aluminum as a disease agent (A1+ through A7+) and evidence against genetic causes (G1- through G3-), while

the other participant (P2) received evidence for genetic causes (G1+ and G2+) and evidence against aluminum (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.

Procedure

After signing of consent forms and a demographic survey, participants were introduced to the software and format of the study sessions through a standardized set of instructions and demonstrations. Participants were then led to their respective stations in different rooms from each other, and worked for up to 30 minutes on a "warm-up" problem to familiarize themselves with the software. Then participants were given up to 120 minutes to work on the main problem, Guam ALS-PD. All six study sessions were completed within this time period.

In order to inform online learning, we designed an asynchronous communication protocol that enabled us to conduct experimental sessions with participants in the laboratory. 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 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 protocol in which (1) participants took occasional "breaks" from their work to play a computer game, Tetris™, and (2) the work of the other participant became available only after these breaks. In each study session, participants were expected to read the four articles and update the shared workspace as they deemed appropriate. Tetris™ was chosen for its familiarity and because it presents a perceptual motor activity quite different from the cognitive task of the experiment, in this sense constituting a break from the primary task.

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 online environment remained available to each participant during the essay writing, but there was no further communication between participants. All participants were able to complete their essays in this time period. One week after the experimental session, each participant was required to complete an online post-test (described below) before payment was sent.

Data collection

Demographic information was collected through a survey and by obtaining Scholastic Aptitude Test (SAT) scores and Grade Point Averages (GPA) (with participants' permission). Process data was collected through two primary means. First, the Morae™ recording software was used to capture the computer screen in digital video. Second, our software was designed to generate complete logs of all the events at each client workstation. Post-session data included the essay and the usability questionnaire elicited immediately after the experimental session, and the post-test elicited one week later.

The post-test was a 20-item 6-choice objective question and answer instrument based on information contained in the ALS-PD articles. 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 distracter 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 in combination suggested a cause of the medical condition. Integrative questions were further divided: *high integration* questions required integration of information presented 5 or more study sessions apart (the "inferential span" of Suthers & Hundhausen, 2003). The test design allows us to separate out evidence for information sharing from evidence of integrative elaboration. 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.

Results

Our analyses addressed *outcomes,* based on content analyses of the essays and scoring of the post-test; and *session processes,* based on quantitative analyses of elaboration on hypotheses. Due to space constraints, ANOVA tables are omitted. Tables may be found in (Suthers et al., in press), although the present paper contains further interpretations. The traditional criterion of α≤0.05 is used for statistics computed to test hypotheses. However, we view probabilities as properties of the data to be reasoned about, not merely as input to a mechanical binary decision procedure (Gigerenzer, 2004). Therefore we report *p* values of 0.1 and below as indicative of phenomena worthy of further investigation.

Outcomes analyses

Content analysis of individually written essays examined both participant's conclusions (disease hypotheses), and the facts participants cited from the information we provided, with particular attention to evidence for sharing of information given to only one participant.

We coded each participant's essay for information that was provided by the source materials. The point of this coding was to trace out ideas that came from the source materials. Therefore, the coding units were based on information as it was expressed in sentences and figures of the source materials. Two analysts independently carried out the analysis and conflicts were resolved by consensus. The counts are presented in Table 2, factored in three ways for the following tests. (1) As a baseline, we wanted to determine whether the treatment groups differed on the amount of information participants collectively expressed in their essays, and on the source of that information (from materials given to P1 versus materials given to P2). We conducted a post-hoc two-way analysis of variance (ANOVA) between the three conditions (Text, Mixed and Graph) and the two sources of materials (given to P1 and to P1), testing for an interaction effect. The dependent variable was the amount of information cited in each participant's individually written essay. The difference was not significant. (2) A follow-up one-way ANOVA did not indicate significant differences across the three conditions on participant's preference for facts from their own materials versus others'. (3) We then wanted to see whether there was a difference between the three conditions in the amount of information cited by *both* P1 and P2 in the individually written essays. The follow-up one-way ANOVA was not significant by the criterion of α≤0.05, but can be interpreted as being consistent with greater overlap in Text ($F(2,27)=2.82$, $p=0.0771$).

Table 2: Mean counts of information units in the essays (columns denote source materials)

(1) Information units in P1 or P2's essays

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

(2) Breakdown of (1) for individual essays

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

(3) Information units in both P1 *and* P2's essays

	From P1	From P2	Total
Text	2.5	3.4	5.9
Graph	2.1	1.6	3.7
Mixed	1.1	2	3.1

We also examined the conclusions 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. Participant's conclusions were assessed on 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 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 3. There are clear differences between treatment groups in pair agreement, with greater convergence in the Graph condition ($\chi^2(2, N=30)=7.5$, $p \leq 0.025$). From the standpoint of quality of solution (under an admittedly simple measure), the difference is decidedly not significant, in spite of the appearance of a trend in the table.

Table 3: Conclusions selected in essays

	Convergence *	Quality
Text	4/10	5/20
Graph	8/10	2/20
Mixed	2/10	2/20

Recall that the *post-test* included both memory and integrative questions. 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 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. 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 ($F(2,57)=4.40$, $p=0.0167$). A Bonferroni 95% CI indicated that Graph participants performed better than Mixed participants.

Process analysis of study session data

Analyses of the study sessions themselves enable us to identify possible explanations for the outcome differences. Although most of these quantitative analyses were planned, exploratory examination of the session logs led to an unplanned quantitative analysis. In the 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. 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.

A post-hoc test of the *time to consider the first hypothesis* 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, measured in seconds from the start of the first ALS-PD study session ($M=618$, $SD=568.9$) The Mixed condition was ranked next ($M=1162$, $SD=1244.3$) as compared to the Text condition ($M=2433$, $SD=1807.7$). A Bonferroni 95% CI showed that the differences lie between Text and Graph, and between Text and Mixed.

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 purposes of this analysis, *elaboration* is defined to include 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 ANOVA of the *total elaborations* on hypotheses revealed significant differences between the groups ($F(2, 57)=13.59$, $p<0.0001$). There were more elaboration acts in the two treatment conditions that offer an evidence mapping tool: both Graph ($M=17.90$, $SD=13.74$) and Mixed ($M = 12.85$, $SD=7.05$) had considerably more elaborative acts than Text ($M=3.25$, $SD=2.45$). A one-way ANOVA of the *number of hypothesis expressed* revealed significant differences between the treatment groups ($F(2, 57)=4.73$ $p=0.0126$). Participants in Graph expressed significantly more hypotheses ($M=5.7$, $SD=3.1$) than in Text ($M = 3.3$, $SD = 1.7$). As would be expected from these results, a one-way ANOVA of the *average number of elaborations per hypothesis* was significant ($F(2, 57)=6.86$, $p<0.0021$). The differences are between both Graph ($M = 3.785$, $SD = 3.634$) and Mixed ($M = 3.781$, $SD = 2.981$) versus Text ($M = 0.995$, $SD = 0.762$): the presence of an evidence mapping tool results in more elaboration on each idea considered.

Discussion

Two lines of evidence support H1, 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 (i.e., in earlier study sessions) and there was more elaboration on the hypotheses individually as well as collectively. Furthermore, Graph users considered more hypotheses. 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 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 number of facts mentioned in the essay (content analysis 1), on whether information given to one participant appeared in the other's essay (content analysis 2), on the information that both participants found worth citing in the essay (content analysis 3), or on memory for information given to one's partner (post-test analysis). 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. An experimental 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, the lack of differences on quality of solution may be counted as evidence against H1. The slightly greater overlap in Text participant's essay content (which did not reach $\alpha \leq 0.05$ in analysis 3) 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. The final set of messages available in the sequential representation might 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, but the dual workspace is a confounding factor, as it requires managing two representations (Ainsworth, Bibby, & Wood, 1998). Participants in the Mixed condition may have converged the least because the dual workspaces provide more variation in strategies for using the workspaces, increasing the possibility that members of a pair will look at different material.

Turning to the comparison between H2 (in favor of integrated representations such as Graph) and H3 (in favor of distinct discussion and conceptual representations such as Mixed), direct differences between Graph and Mixed are limited, the exception being that Graph users remember more integrative relationships than Mixed. Again, the additional complexity of using two representations (the threaded discussion and the evidence map) may have been a factor in Mixed. 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. However, 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. In contrast, Mixed was sometimes advantageous to Text, sometimes not, but never was advantageous to Graph, and sometimes yielded the worst results. Since Graph and Matrix were introduced as competing alternatives to threaded discussions, support for H2 is stronger than for H3.

Summary and Conclusions

Many tools for online collaborative learning are text based, typically providing representational support only for conversational 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 experiment 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 discussion representations. This experiment was undertaken in an asynchronous setting, using a protocol for practical experimental study of asynchronous collaboration in the laboratory. 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 they had considered more hypotheses and appeared to have access to the same information. 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 that traced out information sharing during the session would provide more direct evidence. Such an analysis has recently been completed, and is reported in a companion paper (Suthers, Medina, Vatrapu, & Dwyer, 2007), because it addresses a distinct research question (comparing models of collaboration rather than software conditions) and relies on a different form of analysis.

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 the results point, it would be a brash conclusion to draw from this experiment 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. Any workspace has a limited useful life before it becomes important to "rise above" the clutter and start fresh (Scardamalia, 2004). The subject matter, task structure, and nature of the representations used could also affect results. 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. Clearly, there are ample opportunities for further research in the "middle space between communication and information interfaces" (Hoadley & Enyedy, 1999).

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.

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.

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.

Gigerenzer, G. (2004). Mindless statistics *The Journal of Socio-Economics, 33*, 587-606.

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.

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.

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.

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.

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.

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

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.

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., & 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., Medina, R., Vatrapu, R., & Dwyer, N. (2007). Information Sharing is Incongruous with Collaborative Convergence: The Case for Interaction. In *Computer Supported Collaborative Learning 2007 (this volume)*.

Suthers, D. D., Vatrapu, R., Medina, R., Joseph, S., & Dwyer, N. (in press). Beyond threaded discussion: Representational guidance in asynchronous collaborative learning environments. *Computers & Education, doi:10.1016/j.compedu.2006.10.007*.

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.

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

Acknowledgments