

Vicarious Learning: Effects of Overhearing Dialog and Monologue-
Like Discourse in a Virtual Tutoring Session

David M. Driscoll, Scotty D. Craig, Barry Gholson,
Matthew Ventura, Xiangen Hu, & Arthur C. Graesser

The University of Memphis

Driscoll, D., Craig, S. D., Gholson, B., Ventura, M., Hu, X., & Graesser, A. (2003).
Vicarious learning: Effects of overhearing dialog and monolog-like discourse in a
virtual tutoring session. *Journal of Educational Computing Research*, 29, 431-
450.

Running head: Vicarious Learning

Contact person: Barry Gholson, Department of Psychology, The University of Memphis,

Memphis, TN 38152

Phone: 901 678-4697

Fax: 901 678-2579

<b.gholson@mail.psyc.memphis.edu>

Abstract: In two experiments, students overheard two computer-controlled virtual agents discussing four computer literacy topics in dialog discourse and four in monologue discourse. In Experiment 1, the virtual tutee asked a series of deep questions in the dialog condition, but only one per topic in the monologue condition in both studies. In the dialog conditions of Experiment 2, the virtual tutee asked either deep questions, shallow questions, or made comments. In a fourth “dialog” condition, the comments were spoken by the virtual tutor. The discourse spoken by the virtual tutor was identical in the dialog and monologue conditions, except in the fourth dialog condition. In both studies, learners wrote significantly more content and significantly more relevant content in the deep question condition than in the monologue condition. No other differences were significant. Results were discussed in terms of advanced organizers, schema theory, and discourse comprehension theory.

Vicarious Learning: Effects of Overhearing Dialog and Monologue- Like Discourse in a Virtual Tutoring Session

Recent advances in educational technology, distance learning, and instructional design (Mayer, 2001), have created situations in which learners are more and more likely to find themselves trying to understand talk in settings in which they are over-hearers (Fox Tree, 1999; Schober & Clark, 1989), rather than active participants. The new educational technologies stimulate a host of questions for researchers in cognition and instruction. One of these questions concerns the need for further empirical understanding of the conditions under which over-hearers can benefit in language environments in which they are relatively isolated.

In the research reported below, we adopted a vicarious-learning perspective (Bandura, 1962; McKendree, Stenning, Mayes, Lee, & Cox, 1998). Historically, the term vicarious learning was frequently used synonymously with social learning, or observational learning (Bandura, 1977). According to this perspective, through simply observing activities carried out by others, learners can master how to perform the activities without overt practice or direct incentives (Rosenthal & Zimmerman, 1978, p. xi). In the present context the question addressed is: “What benefits can students gain from dialogue as observers, not just as participants?” (Lee, Dineen, & McKendree, 1998, p. 17).

Craig, Gholson, Ventura, Graesser, and TRG (2000) previously reported research designed to explore whether vicarious-learning procedures could be used effectively to induce deep-level question asking (Bloom, 1956; Graesser & Person, 1994) in a transfer task. A computer-controlled virtual tutor and a virtual tutee, located on a monitor,

discussed a series of eight computer-literacy topics. During acquisition, learners either overheard the virtual tutee carry on a dialog with the virtual tutor, or they overheard a more monologue-like discourse. Each topic began with a brief information delivery by the virtual tutor. Then, in the monologue-like condition, the virtual tutee asked one broad question relevant to the topic, and the virtual tutor answered in a monologue that presented all the information on that topic. In the dialog condition, each information delivery was followed by a series of conversational exchanges. The virtual tutee asked a series of specific questions on each topic and the virtual tutor immediately answered each. The tutorial content was identical in the two conditions. Free-recall questions on the discourse contents were administered following both acquisition and transfer.

In the transfer task the learners were presented with a series of eight new topics and were given the opportunity to ask questions on each. The prediction was that if the vicarious dialog condition used in acquisition was successful in inducing question asking in the transfer task, the students in that condition would ask more questions, and consequently, they would learn more as reflected in their performance in the free-recall protocols. The Craig et al. study confirmed the predictions. In the transfer task learners in the dialog condition took significantly more conversational turns, generated significantly more queries, and asked significantly more deep-level questions (Craig et al., 2000, pp. 248-249) than those in the monologue-like condition. In the free-recall task following transfer, students in the dialog condition also wrote significantly more contents than those in the monologue-like condition, with an effect size of .53 (Cohen's d). An unexpected finding was that learners in the dialog condition also provided more written contents in the free-recall questions following acquisition than those in the monologue-like

condition. This latter finding was not significant, but Cohen's d yielded an effect size of .44. This serendipitous result suggested a question that has not been previously addressed: Do students learn (i.e., recall) more when they overhear dialog discourse than when they overhear monologue-like discourse containing the same expository contents?

There is at least some evidence that they might. Fox Tree (1999) reported that over-hearers performed better in a referential communication task while overhearing dialogues than while overhearing monologues. In a preliminary task, Fox Tree divided students into pairs, called Directors and Matchers (Schober & Clark, 1989). The goal was for the Director to describe an ordered set of abstract shapes (tangrams) to the Matcher so the Matcher could place the shapes in the same order as the Director's. Directors either gave instructions for the Matcher to follow (monologue condition), or conversed freely with the Matcher (dialogue condition). The sessions were taped, and only those sessions in which Matchers correctly ordered the tangrams were used in the vicarious matching experiment. As noted above, the results were that the vicarious matchers who overheard dialogs outperformed those who overheard monologues.

Experiment 1

From the Craig et al. (2000) study, we know that overhearing question-rich dialog promotes learners' question asking in a transfer task, and that the question asking during transfer promotes learning. What remains to be determined is whether overhearing a question-rich dialog promotes learning of the overheard contents. Thus, Experiment 1 was designed to explore whether the non-significant trend in favor of those who overheard dialog during acquisition in the Craig et al. (2000) study would prove more robust if more precise measures were used. In the Craig et al. study, discourse type (i.e.,

monologue vs. dialog) was a between-subject manipulation, and the data were quite variable. For example, some learners wrote only one or two brief sentences on each free-recall question, while others wrote more than a (8 1/2- x 11-in) page on each. Thus, in order to obtain more precise measures of the learners' performances, discourse type was a within-subjects variable in the present experiments. Similar to the Craig et al. study, a computer-controlled virtual tutor and virtual tutee engaged in discourse concerning eight computer-literacy topics. Each learner overheard four computer literacy topics discussed in dialog format and four in monologue-like format. Following this exposure, each participant was given two free-recall test questions, one probing a topic overheard in the dialog format and the other probing a topic overheard in the monologue-like format.

Method

Participants and Design

A total of 48 students, whose participation met a course requirement, were drawn from introductory psychology classes at the University of Memphis. Because earlier research has demonstrated that learning effects are stronger among students with low domain knowledge (e.g., Craig, Gholson, & Driscoll, 2002; Mayer, 2001; Moreno & Mayer, 1999), only students who scored six or less on a 12-item knowledge questionnaire adapted from Mayer and Moreno (1998, p. 34) were included in the study. The design of the experiment was a 2 (discourse type: monologue vs. dialogue) x 2 (tutorial content: software vs. hardware) factorial. We included tutorial content as a between-subject variable in order to assure that any significant findings did not result from specific tutorial contents. Due to the large between-subject variability described in the previous section, discourse type was a within-subject variable. Discourse type was

counterbalanced across participants. That is, half of the students exposed to each type of tutorial content (i.e., hardware or software) over-heard four topics discussed in dialog format followed by four topics discussed in monologue-like format, while the other half over-heard four topics in the monologue-like format followed by four in dialog format. The 48 participants were randomly assigned to the resulting four conditions, with 12 per group. The data of four students were lost due to failure to follow instructions: Two of these were students who overheard dialogs followed by monologues, and the other two were in the other counterbalancing condition. Students were tested individually in one session lasting about 40 minutes.

Materials and Procedures

The computerized materials were created using two computer applications, Xtrain (Hu, 1998) and Microsoft Agent (Microsoft, 1998). Xtrain was created for the purpose of integrating Microsoft agents with text, pictures, HTML, Macromedia Flash, and Director files. It was used to script and control the presentations of the virtual agents, narration, and the picture displays. Microsoft Agent (1998) was used for character creation, animation, and voice synthesis of the virtual tutor and virtual tutee.

Tutorial content for topics concerned with both software and hardware was derived from curriculum scripts for an intelligent tutoring system (ITS). This ITS was created by the Tutoring Research Group (TRG) at the University of Memphis (Graesser, Wiemer-Hastings, K., Wiemer-Hastings, Kreuz, & TRG, 1999; Person, Graesser, Kreuz, Pomeroy, & TRG, 2001). Half the participants were presented with a virtual tutoring session involving eight hardware topics; for the remainder, the content was eight software topics. Each student was presented with dialog discourse on four of the topics, and

monologue-like discourse on the other four. A total of 66 questions were constructed for use in the dialog discourse for each type of tutorial content. Learners assigned to dialog on the first four topics overheard 33 questions, and learners assigned to dialog on the second four topics overheard the other 33. One question was constructed for each topic in the monologue discourse. In both the monologue and dialogue discourse, each topic was introduced by a brief information delivery presented by the virtual tutor. In the monologue-like discourse, the virtual tutee then asked one rather broad question, designed to justify the monologue that followed, and the virtual tutor followed this with a monologue presentation of all the tutorial content for that topic. In the dialogue discourse, after each information delivery the virtual tutee asked a series of rather specific questions. The exact words, phrases, and sentences spoken by the virtual tutor were identical in both the dialogue and monologue-like discourse in each content domain (i.e., software and hardware).

The virtual tutor, the virtual tutee, and a picture relevant to the particular topic under discussion were displayed on the monitor at all times during the session. The picture changed from topic to topic in both content domains. A screen shot, showing the two animated agents and a picture used for a topic on computer speed is presented in Figure 1. This picture depicts two CPUs, one with a 16-bit buss and the other with a 32-bit buss. The latter simultaneously carries twice as much information as the former. The topic was concerned with how manufacturers can increase the speed of computers. At the end of a topic the virtual tutor indicated that they (virtual tutor and virtual tutee) were going on to a new topic and brought up a new picture.

Immediately following the eighth topic, each student was given a retention test that consisted of two open-ended questions, each probing one of the topics covered by the virtual tutor and virtual tutee. One free-recall question presented to each student probed a topic that was covered in the dialog format between the virtual tutor and tutee, and the other probed a topic that was discussed in the monologue-like format. The free-recall questions were administered in the same order as discourse type. Consequently, for half the students a topic overheard in dialog discourse was probed first; for the other half the first topic probed was discussed in monologue-like discourse.

Insert Figure 1 about here

An example retention question is presented in Appendix A. The questions were administered one at a time, and students were permitted to write until they said they were finished. For purposes of assigning retention-test questions to learners, constrained randomizations were prepared. This procedure was followed in order to assure that each topic presented in dialog format and each presented in monologue format was probed an equal number of times in each of the four groups. Among the 12 students assigned to each subgroup (e.g., hardware content with dialog followed by monologue), each topic question was presented to three different students, and no two students were tested on exactly the same two topics. This was done in an attempt to assure that outlier performance of a few students would not bias the free-recall data.

Questions asked by the virtual tutee in both the monologue and dialog discourse reflected the naïve perspective of an inquisitive novice on first exposure to the contents of the computer-literacy topics. The virtual tutor's perspective was always that of an accomplished expert. Because the virtual tutee had only one conversational turn in the

monologue-like condition on each topic, and it was intended to justify the lengthy monologue that followed, questions in that discourse condition were generally more comprehensive than those in the dialogue condition. For example, in a topic on the operating system, in the monologue-like discourse the question asked by the virtual tutee was: “When you first turn on a computer they say it boots up. What I would like to know is what booting up means? What is actually going on right after you first turn on a computer?” In the corresponding dialogue discourse, following the information delivery the first question asked was “Why does it load the operating system, what does that accomplish?”.

Results and Discussion

For purposes of analysis, data from the free-recall questions were scored by computing the total number of propositions (i.e., story units) that were written by learners on each of the two free-recall questions they were presented. In order to standardize the scoring, procedures described by Bovair and Kieras (1985) were used, and each proposition included one predicate and exactly two arguments. Their methods emphasize representing the prepositional content in the simplest way possible. For example, tense is disregarded, auxiliaries are not represented, it avoids embedding when possible, and when a verb is followed by a preposition, the preposition is treated as part of the verb itself. The story units written by each learner were then classified into three categories: relevant, related, and irrelevant. To accomplish this, the contents spoken by the virtual tutor on all 16 topics (eight hardware, eight software) were first propositionalized using the scoring procedure described directly above. Propositions written by each learner on the free-recall tests were then compared to those spoken by the virtual tutor.

Propositions written by learners on a given topic that matched (or paraphrased) those spoken by the virtual tutor on that topic were classified as relevant. Consider an example of paraphrase on a topic concerned with computer speed (Figure 1). The virtual tutor said “A wider buss transmits more information.” This was represented as two propositions: (TRANSMITTS, WIDER BUSS, INFORMATION), and (TRANSMITTS, MORE, INFORMATION). An acceptable paraphrase would be “A bigger buss carries more data,” represented as (CARRIES, BIGGER BUSS, DATA), and (CARRIES, MORE, DATA). Any proposition that matched (or paraphrased) the contents spoken by the virtual tutor on any of the other topics was classified as related to the contents of the topic being scored. Any other propositions written by the learner were classified as irrelevant. Examples included mostly side comments, such as “I don’t know much about computers;” or (occasional) propositions concerning the tutorial content that were false. All the data were classified by two of the authors: Pearson correlations between the two yielded $r = .97$ for relevant information, $r = .93$ for related information, and $r = .98$ for the irrelevant category. The mean number of propositions and standard deviations in each category for the dialog and monologue-like conditions are presented in Table 1.

Insert Table 1 about here

Because discourse type was counterbalanced, a preliminary analysis was performed to assess order effects. A one-way analysis of variance failed to yield significance, $F(1,42) = 1.37$, $p > .05$, and order was excluded from further analyses. A 2 (discourse type: dialogue vs. monologue) x 2 (tutorial content: software vs. hardware) analysis of variance performed on the irrelevant propositions revealed no significant

effects, and the means were roughly comparable (dialogue \underline{M} = 2.48, monologue \underline{M} = 2.36). A similar analysis performed on related propositions also failed to reveal any significant differences (dialog \underline{M} = 2.25, monologue \underline{M} = 2.23). Analysis of the relevant propositions revealed only a significant effect of discourse type, $F(1,42) = 27.49$, $p < .001$. Learners wrote significantly more relevant information on free-recall questions probing topics overheard in dialog format (\underline{M} = 15.41) than they wrote on questions probing topics overheard in monologue-like format (\underline{M} = 8.82). Cohen's d statistic yielded an effect size of .71. Clearly, then, the students learned more of the tutorial content when it was overheard in dialog format than when the same content was overheard in a monologue-like format. This finding has potentially important implications for those who design electronic educational environments for students who are more likely to be involved as onlookers and over-hearers than as active participants. What remains to be determined, of course, is what specific features of dialog discourse are important when it is compared to more monologue-like discourse?

Experiment 2

Experiment 2 was designed to explore four features of dialog that might have accounted for the differences obtained in Experiment 1. It was deemed possible that the virtual tutee's contributions to the dialog might have improved vicarious learning (a) because they functioned as signaling devices similar to headings in printed text, (b) because they provided concept repetition, (c) because the contributions were questions, and/or (d) because they were questions that provided a coherent context for the information spoken by the virtual tutor on his next turn. Each of these four possibilities is discussed in turn.

First, it is well known that including headings in printed text as advanced organizers can facilitate text-processing strategies and recall (e.g., Corkill, 1992; Hartley & Trueman, 1985; Lorch & Lorch, 1995). Thus, the tutee's contributions to the dialog discourse in Experiment 1 may have functioned as signaling devices (Lorch & Lorch, 1996), in a manner similar to the role of headings in printed text.

A second possibility, which Fox Tree (1999) suggested might be an important consideration when dialog is compared to monologue, involves concept repetition. Most of the virtual tutee's contributions in the dialog condition of Experiment 1 queried key concepts mentioned by the virtual tutor in his previous conversational turn. In addition, the question that was asked usually anticipated the contents elaborated on in the virtual tutor's next conversational turn. Thus, even though the tutee, in a sense, only named the concepts, and did not provide explanatory content about them, the repetition itself might have facilitated performance in the dialog condition.

A third possibility concerns the form of the virtual tutee's contributions to the dialog. In Experiment 1, the virtual tutee asked a total of 33 questions across the four topics covered in dialog discourse, but only one question per topic on the four covered in monologue-like discourse. This question-rich feature of the dialog may have been important, because there is some evidence that questions *per se* play an important role in vicarious learning (McKendree, Good, & Lee, 2001). McKendree et al. had over-hearers listen to dialogs taken from an existing Map Task Corpus (see Map Task Corpus, below in References). In the original map task (Brown, Anderson, Shillcock, & Yule, 1984), a Giver and a Follower were each given a schematic map with a number of landmark features. One of the pair, the Giver, had a route from a start point to an end point drawn

on the map. The task was for the Givers to describe the route drawn on their maps to the Followers, who attempted to draw in the route from the start point to the end point. Most of the landmark features were shown on both versions of the map, but in order to complicate the task, there were some discrepancies. Some features were shown only on the Givers' maps and some were shown only on the Followers' maps. In addition, some features were named differently on the Givers' and Followers' maps. These differences were systematically manipulated in the original corpus (Brown et al., 1984) in order to explore the impact of shared information on the resulting dialogs. Dialogs from the map task were recorded as Givers and Followers discussed drawing in the route.

McKendree et al. (2001) presented over-hearers with recordings of one or another of two dialogs taken from the Map Task Corpus. Both dialogs were taken from the same condition in the original corpus. There were five discrepancies between the Givers and the Followers maps in that condition. Although the two dialogs were equated for length, one dialog included about twice as many questions as the other dialog. Over-hearers were given maps like those used by Followers in the condition from which the dialogs were taken. Their task was to draw in the route from the start point to the end point, while overhearing one of the dialogs. What McKendree et al. found was that over-hearers who listened to the question-rich dialog were able to draw much more accurate routes (using map discrepancy scores) than those who overheard the dialog discourse that contained few questions. They (McKendree et al., 2001) concluded that the use of questions as part of the overheard dialog was a critical feature that promoted vicarious learning. This suggests the possibility that simply overhearing a large number of questions may have facilitated performance in the dialog condition of Experiment 1.

A fourth feature of the discourse that is explored in the current experiment concerns the kinds of questions that were asked by the virtual tutee. The questions from the dialog condition of Experiment 1 were classified according to the Graesser, Person, and Huber (1992; Craig et al., 2000; Graesser & Person, 1994) question taxonomy. This taxonomy contains five short-answer categories and 11 categories requiring long answers. Nearly all of the questions used in Experiment 1 were drawn from long-answer categories. These latter question categories have been shown to correlate reasonably well ($r = .64$) with some of the deeper-level reasoning questions (levels 2-5) of Bloom's (1956) taxonomy (Graesser & Person, 1994, p. 128). These levels include comprehension, application, analysis, and synthesis, as opposed to level 1, which involves only specific bits of information (Bloom, 1956). Thus, it is possible that it was the type of questions asked by the virtual tutee, rather than questions per se, that was responsible for the enhanced recall. If not, and questions per se are a critical feature of dialog, as suggested by McKendree et al. (2001), then short-answer (shallow) questions should be as effective in promoting vicarious learning from overheard dialog as (deep reasoning) long-answer questions.

If deep questions do promote vicarious learning from overheard discourse, what mechanism is responsible? One possibility concerns the coherence of the information the learners were asked to comprehend. According to Gernsbacher's (1990, 1997) theory, called the "structure building framework," the goal of comprehension is to build coherent mental representations or mental structures. The theory says that coherence arises from at least four sources: referential, temporal, locational, and causal continuity (Gernsbacher, 1997, p. 269). After preliminary information lays the foundation for the mental structure,

the structure is built by mapping new information that coheres or relates to the previous information. If the new information is less coherent, or is not closely related to the foundation, a new and different substructure is built. The idea is that causally coherent information is easier to map to the existing mental structure than less coherent or unrelated information (Gernsbacher, 1997, p. 270), and the causal coherence makes the information easier to process and recall. In Experiment 1, most of the questions asked by the tutee in the dialog condition involved either (causal) enablement, causal consequence, or causal antecedent (Graesser & Person, 1994), and discourse researchers have shown that causal connectives facilitate both reading times and recall (e.g., Graesser, Millis, & Zwaan, 1997; Murray, 1995). Thus, discourse theory, and Gernsbacher's (e.g., 1997) structure building framework in particular, suggest that the series of deep-reasoning questions asked by the virtual tutee on each topic in Experiment 1 may have provided a coherent causal context for the informational content spoken by the virtual tutor. This context would be expected to improve recall by making the incoming content more causally coherent, easier to process, and easier for the vicarious learner to map on to the growing mental structure.

Another possibility involves cognitive schema theory (Sweller & Cooper, 1985; Derry, 1996). According to Sweller and Cooper, even low-level schemas permit people to recognize patterns so they can respond with appropriate actions. These kinds of schemas represent basic knowledge, but they are structured and complex. A higher type of schema is called an object family. They involve loosely organized collections of knowledge that work together in given situations. Objects of knowledge within an object family activate each other and behave as a single schema. Another type of schema, that involves a

distributed type of memory activation, is called a cognitive field (Derry, 1996). Some patterns of memory activation are made more accessible than others when these schemes are activated by particular events (Bransford & Johnson, 1972). So when these schemas are activated in a learning situation they determine how previously-existing knowledge can be modified by new information. It is possible that the deep-level questions used in Experiment 1 activated schemas of various sorts, and that this activation made incoming content easier to process and map to the existing knowledge base.

In order to explore the various possibilities discussed above, Experiment 2 included four dialog conditions as a between-subjects variable, with discourse type (i.e., dialog vs. monologue) as a within-subject manipulation in each. As in Experiment 1, learners overheard four computer literacy topics discussed in dialog discourse and four in monologue-like discourse. In one condition, the virtual tutee's contributions to the dialog were all deep (long-answer) questions (Graesser & Person, 1994) and in a second they were all shallow (short-answer, yes/no) questions. In a third condition the virtual tutee's contributions were all assertions that were transformations of the shallow questions. These assertions were spoken by the virtual tutee in what is called the "comment" condition. An excerpt taken from a topic concerned with computer speed is presented in Appendix B. This illustration includes an example information delivery, called "Virtual tutor information delivery," along with the virtual tutee's contributions in the monologue-like condition, labeled "Virtual tutee monologue question condition," and the virtual tutee's contributions in the three dialog conditions. The latter are called the "Virtual tutee dialog deep question condition," "Virtual tutee dialog shallow question condition," and "Virtual tutee dialog comment condition" in Appendix B. The virtual tutor's

contributions after the information delivery are labeled “Virtual tutor all conditions.” In a fourth “dialog” condition, the virtual tutor’s spoken contributions included both those illustrated by “Virtual tutor all conditions.” and that provided by the virtual tutee in the comment condition. This is referred to as the “enhanced monologue” condition in further discussion.

If concept repetition, by itself, enhances vicarious learning, performance in all four of the dialog conditions should exceed that exhibited in the monologue condition. If the tutee’s contributions to the dialog function as signals, similar to the role of headings in printed text, then the deep-question, shallow-question, and comment conditions should all produce differences when the dialog conditions are compared with the monologue condition, but the enhanced monologue condition should not. If questions *per se* facilitate learning from overheard dialog (McKendree et al., 2001), then both the long-answer (deep reasoning) and short-answer (shallow) question conditions should produce differences when the dialog conditions are compared to the monologue-like condition, while the other two conditions should not. Finally, if deep-reasoning questions were the key feature of the overheard dialog in Experiment 1, then only learners overhearing the long-answer questions would be expected to show significantly enhanced vicarious learning when the dialog condition is compared to the monologue-like condition.

Method

Participants and Design

A total of 96 students, whose participation met a course requirement, were drawn from introductory psychology classes. The checklist from Experiment 1 was used to screen potential participants. Only those who reported they had no more than average

knowledge of computers were included. The basic design was a 2 (discourse type: dialog vs. monologue) x 4 (dialog condition: deep question vs. shallow question vs. comment vs. enhanced monologue) factorial. Discourse type was a within-subject manipulation, while dialog condition was a between-subject factor. Discourse order was counterbalanced across participants, with half the participants in each group overhearing four topics discussed in monologue-like format followed by four in dialog format, and for the other half overhearing the discourse types in the reverse order. The 96 participants were randomly assigned to the resulting eight groups, with 12 in each. Experimental sessions lasted about 40 minutes.

Materials and Procedures

The computerized materials were created using the same software as in Experiment 1 (Hu, 1998; Microsoft, 1998). Each learner again overheard a computer-controlled virtual tutor and virtual tutee, located on a monitor, discuss four computer-literacy topics in monologue-like discourse and four topics in one of the dialog conditions, with a picture relevant to the topic under discussion also located on the monitor. Each topic was introduced by a brief information delivery presented by the virtual tutor. Because there was no hint of an effect of tutorial content (hardware vs. software) in Experiment 1, or in an earlier study (Craig et al., 2000), only hardware topics derived from curriculum scripts created by the TRG (see Experiment 1, above) were used in Experiment 2.

The virtual tutee's style of discourse with the virtual tutor varied across the four dialog conditions. The exact words, phrases, and sentences spoken by the virtual tutor were, however, identical across the deep-question, shallow-question, and comment dialog

conditions, as well as in the monologue-like condition on each topic. For the deep-question condition, the few questions that required short answers in Experiment 1 were modified to require long answers. Thus, in the deep-question condition, all 33 of the questions assigned to both counterbalancing conditions involved long answers. That is, all 33 questions overheard by learners assigned to the dialog condition for the first four topics, and all 33 questions overheard by those assigned to the dialog condition for the second four topics involved long answers. These questions presumably reflected deep reasoning (e.g. “How can the hardware keep track of things if the switches have only two settings?”), and encouraged over-hearers to encode information spoken by the virtual tutor in terms of causally coherent relations (Gernsbacher, 1997), or activated schemas that made incoming information easier to process (Derry, 1996). In the shallow-question condition, all 33 of the questions overheard by learners in each counterbalancing condition required short, one-word (yes/no) answers and presumably reflected shallow reasoning (e.g., “Are two settings enough to keep track of everything?”) that did not promote encoding in terms of causal relations. In the comment condition, the short answer questions were transformed into simple assertions (see Appendix B) spoken by the virtual tutee (e.g., “Two settings are enough to keep track of everything.”).

In a final “dialog” condition, dubbed “enhanced monologue,” the simple assertions spoken by the tutee in the virtual tutee dialog comment condition were instead inserted into the discourse spoken by the virtual tutor. That is, in the enhanced-monologue condition, after the virtual tutee asked the monologue question the virtual tutor spoke both the contents illustrated by “Virtual tutee dialog comment condition” and those illustrated by “Virtual tutor all conditions” in Appendix B. In the standard

monologue-like condition, following the information delivery by the tutor, the virtual tutee asked the one question (e.g., see “Virtual tutee monologue question condition” in Appendix B), and it was followed only by the virtual tutor’s monologue. The monologue included only those contributions labeled “Virtual tutor all conditions,” as illustrated in Appendix B.

After overhearing all eight hardware topics, each participant answered two free-recall questions (e.g., Appendix A). One question probed a topic overheard in dialog format, and the other queried a topic overheard in monologue-like format. The questions, presented one at a time, were counterbalanced, with each learner receiving the questions in the same order as they overheard the discourse (i.e., dialog vs. monologue). That is, learners who overheard dialog discourse followed by monologue-like discourse were presented with a free-recall question probing a dialog topic first, while those who overheard monologue-like discourse first received a question probing a monologue topic first. Across the 12 participants in each subgroup, each free-recall question was presented to three different students, and no two students were tested on exactly the same two topics.

Results

For purposes of analysis, data from the free recall-questions were scored by one of the authors by first computing the total number of propositions written by learners on each of the two free-recall questions. In order to standardize the scoring procedure (Bovair & Kieras, 1985), each proposition included one predicate and exactly two arguments, as in Experiment 1. Free-recall questions from 20 participants were randomly selected and scored by a second author. Agreement was good, yielding a Pearson $r = .98$.

The propositions written by each learner were then classified by one of the authors into three categories: relevant, related, and irrelevant. The free-recall protocols of 20 participants, those scored as containing the greatest mix of relevant, related, and irrelevant propositions, were then scored by a second author. Agreement was reasonably good: relevant $r = .89$, related $r = .92$, irrelevant $r = .90$. The means and standard deviations of the numbers of relevant, related, and irrelevant propositions for each group are presented in Table 2.

Insert Table 2 about here

As in Experiment 1, discourse type was counterbalanced, so a preliminary analysis was performed to evaluate order effects. A one-way analysis of variance failed to yield significant effects, $F(1,94) = 2.75$, $p > .05$, and order was excluded from further analyses. A 2 (discourse type: dialog vs. monologue) x 4 (dialog condition: deep question vs. shallow question vs. comment vs. enhanced monologue) performed on the irrelevant propositions (see Table 2) failed to yield any significant effects (largest $F(3,92) = .64$, $p > .05$). A similar analysis performed on the related propositions also revealed no significant effects (largest $F(3,92) = 1.45$, $p > .05$). A 2 (discourse type: dialog vs. monologue) x 4 (dialog condition: deep question vs. shallow question vs. comment vs. enhanced monologue) performed on the relevant propositions revealed a significant interaction between discourse type and dialog condition, $F(3,92) = 2.78$, $p < .05$.

Simple main effect procedures performed on the propositions written on topics discussed in dialog vs. monologue-like discourse in each dialog condition yielded a significant difference only in the deep-question condition, $F(1,23) = 9.15$, $p < .01$. The mean number of relevant propositions written on topics overheard in the dialog condition

was 17.33, while in the monologue condition the mean number was 11.37 (see Table 2). The effect size was .64 (Cohen's d statistic), which compared favorably with the .71 that was obtained in Experiment 1. Discourse type did not reach significance in analysis of any of the remaining dialog conditions. The largest $F(1,23) = 1.65$, $p > .05$, was obtained in the enhanced-monologue condition. When the four dialog conditions were contrasted with each other (deep question vs. shallow question vs. comment vs. enhanced monologue) in each discourse type taken separately, nothing approached significance. These latter findings reflected the high between-subject variability described earlier. That is, some learners wrote only one or two short sentences on each free-recall question, while others wrote more than a page.

Discussion and Conclusions

Among the four discourse features deemed possibly responsible for the results of Experiment 1, the results of Experiment 2 failed to support three of them. First, concept repetition (Fox Tree, 1999) apparently does not facilitate vicarious learning. If it did we would have expected significant differences between the dialog discourse and monologue-like discourse in each dialog condition. That is, in all four dialog conditions there was repetitious naming of key concepts relative to the monologue-like condition. Second, research on advance organizers (e.g., Ausubel, 1960; Mayer, 1979) suggests that, when properly constructed, organizers like printed headings can facilitate learning and retention of text materials, particularly when the text is poorly organized or when learners have low domain knowledge. However, it does not appear that the virtual tutee's contributions to the dialog functioned as signaling devices, similar to the role of headings in printed text (Lorch & Lorch, 1995, 1996). If they had, there should have been

differences between the dialog and monologue-like discourse in three dialog conditions: deep-question, shallow-question, and comment. One possibility is that the deep questions served as more effective advance organizers than either the shallow questions or the comments, but no research could be located that supports this explanation.

As suggested earlier, it is possible that deep questions activate various kinds of cognitive schemas (Derry, 1996; Sweller & Cooper, 1985), and that this activation makes it easier to process incoming information and map it to existing knowledge. It should be pointed out, though, that participants in both Experiment 1 and 2 possessed low levels of domain knowledge prior to their participation in the session. Thus, these learners would presumably have not activated schemas associated with object families or cognitive fields. It is also not apparent why shallow questions or comments would not activate schemas. However, we clearly cannot rule out the possibility that cognitive schema theory could provide an account of our findings.

Another possibility considered earlier is that the deep questions may have provided contextual and causal coherence for the information spoken by the virtual tutor. Gernsbacher's (1990, 1997) structure building theory says that once preliminary information lays the foundation for a mental structure, the structure is built by mapping new information that is coherent with the foundation (Gernsbacher, 1997, p. 269). While the four types of coherence (referential, temporal, locational, causal) are interrelated, we emphasize causal coherence because the questions asked by the virtual tutor in the deep-question condition were designed to support those relations (see Appendix B). There is a considerable amount of evidence that causal coherence encourages mapping (e.g., Bloom, Fletcher, van den Broek, Reitz, & Shapiro, 1990; Gernsbacher, 1997; Graesser et al.,

1997; Keenan, Baillet, & Brown, 1984). As is illustrated in Appendix B, most of the exchanges between the virtual tutee and virtual tutor in the deep-question condition involved the tutee providing a causal context for the tutor's next contribution. Presumably, this had the effect of preparing the over-hearer for the tutor's next contribution to the dialog, adding causal coherence to the incoming information, and allowing it to be mapped to the growing mental structure. Gernsbacher's (1997) theory, then, appears to provide a reasonable, if preliminary, explanation for the present findings. An intriguing question, one that will have to await further research, concerns the source of the deep questions. That is, is a deep-question-rich dialog necessary to enhance vicarious learning, or could a monologue--in which a virtual tutor both asks and answers deep questions--be just as effective in promoting vicarious learning?

In concluding, we need to highlight some limitations of the present research. One limitation concerns the generality of the findings. Would similar results be obtained in other domains, such as scientific systems that involve causally linked sequences of events (e.g., Mayer, 2001)? It is possible that such event sequences provide their own causal coherence (Sweller, 1999). Second, the participants were specifically selected because they lacked domain knowledge. This was because earlier research has demonstrated that learning effects are stronger among students with low domain knowledge (e.g., Craig et al., 2002; Mayer, 2001; Moreno & Mayer, 1999). It is not clear that similar results would be obtained among knowledgeable learners mastering more advanced concepts (e.g., Kalyuga, Ayres, Chandler, & Sweller, in press).

Finally, a more pressing concern involves the dependent measure (i.e., free recall) that was used in both the Craig et al. (2000) study and in Experiments 1 and 2. The

original finding, the trend in the data following acquisition in the Craig et al. (2000) study, was serendipitous, but worth exploring. As indicated earlier, the between-subject variability was very high on the free-recall measure, and may have led to type II errors. So the cautionary note here is that the present research might have revealed differences between dialog and monologue-like discourse in other dialog conditions, if more precise dependent measures had been used. It is also possible that other measures would have revealed significant differences between some of the four dialog conditions. Future research, then, might fruitfully employ pretest/posttest designs and use more precise dependent measures.

References

Ausubel, D.P. (1960). The use of advance organizers in the learning and retention of meaningful verbal material. Journal of Educational Psychology, 51, 267-272.

Bandura, A. (1962). Social learning through imitation. In M. R. Jones (Ed.), Nebraska symposium of motivation (pp. 211-269). Lincoln: University of Nebraska Press.

Bandura, A. (1977). Social learning theory. Englewood Cliffs, NJ: Prentice Hall.

Bloom, B. S. (Ed.). (1956). Taxonomy of educational objectives. Handbook 1: Cognitive domain. New York: McKay.

Bloom, C. P., Fletcher, C. R., van den Broek, P., Reitz, L., & Shapiro, B. P. (1990). An online assessment of causal reasoning during comprehension. Memory & Cognition, 18, 65-71.

Bovair, S., & Kieras, D. E. (1985). A guide to propositional analysis for research on technical prose. In B. Britton & J. B. Black (Eds.), Understanding expository text (pp. 315-362). Hillsdale, NJ: Erlbaum.

Bransford, J. D., & Johnson, M. K. (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall. Journal of Verbal Learning and Verbal Behavior, 11, 717-726.

Brown, G., Anderson, A. H., Shillcock, R., & Yule, G. (1984). Teaching talk. Cambridge: Cambridge University Press.

Corkill, A. (1992). Advance Organizers: Facilitators of recall. Educational Psychology Review, 4, 33-67.

Craig, S. D., Gholson, B., & Driscoll, D. M. (2002). Animated pedagogical agents in multimedia educational environments: Effects of agent properties, picture features, and redundancy. Journal of Educational Psychology, *94*, 428-434.

Craig, S. D., Gholson, B., Ventura, M., Graesser, A. C., & TRG. (2000). Overhearing dialogues and monologues in virtual tutoring sessions: Effects on questioning and vicarious learning. International Journal of Artificial Intelligence in Education, *11*, 242-253.

Derry, S. D. (1996). Cognitive schema theory in the constructivist debate. Educational Psychologist, *31*, 163-174.

Fox Tree, J. E. (1999). Listening in on monologues and dialogues. Discourse Processes, *27*, 35-53.

Gernsbacher, M. A. (1990). Language comprehension as structure building. Hillsdale, NJ: Erlbaum.

Gernsbacher, M. A. (1997). Two decades of structure building. Discourse Processes, *23*, 265-304.

Graesser, A. C., Millis, K. K., & Zwaan, R. A. (1997). Discourse comprehension. In J. T. Spence, J. M. Darley, & D. J. Foss (Eds.), Annual Review of Psychology (pp. 163-189). Palo Alto, CA: Annual Reviews.

Graesser, A. C., & Person, N. K. (1994). Question asking during tutoring. American Educational Research Journal, *31*, 104-137.

Graesser, A. C., Person, N., & Huber, J.D.(1992). Mechanisms that generate questions. In T. E. Lauer, E. Peacock, & A. C. Graesser (Eds.), Questions and information systems (pp. 167-187). Hillsdale, NJ: Erlbaum.

Graesser, A. C., Wiemer-Hastings, K., Wiemer-Hastings, P., Kreuz, R., & the Tutoring Research Group. (1999). AutoTutor: A simulation of a human tutor. Journal of Cognitive Systems Research, 1, 35-51.

Hartley, J., & Trueman, M. (1985). A research strategy for text designers: The role of headings. Instructional Science, 14, 99-155.

Hu, X. (1998). Xtrain 1.0 [computer software]. The University of Memphis: Advanced Learning Technologies Laboratory.

Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (in press). The expertise reversal effect. Educational Psychologist.

Keenan, J. M., Baillet, S. D., & Brown, P. (1984). The effects of causal cohesion on comprehension and memory. Journal of Verbal Learning and Verbal Behavior, 23, 115-126.

Lee, J., Dineen, F., & McKendree, J. (1998). Supporting student discussions: It isn't just talk. Education and Information Technologies, 3, 17-29.

Lorch, R. F., Jr., & Lorch, E. P. (1995). Effects of organizational signals on text processing strategies. Journal of Educational Psychology, 87, 537-544.

Lorch, R. F., Jr., & Lorch, E. P. (1996). Effects of organizational signals on free recall of expository text. Journal of Educational Psychology, 88, 38-48.

Map Task Corpus: www.hcrc.ed.ac.uk/dialogue/maptask.html/

Mayer, R.E. (1979). Twenty years of research on advance organizers. Instructional Science, 8, 133-167.

Mayer, R. E. (2001). Multimedia learning. New York: Cambridge University Press.

Mayer, R. E., & Moreno, R. (1998). A split-attention effect in multimedia learning: Evidence for dual processing systems in working memory. Journal of Educational Psychology, 90, 312-320.

McKendree, J., Good, J., & Lee, J. (August, 2001). Effects of dialogue characteristics on performance of overhearers. Cognitive Science Society Meetings, Edenborough, Scotland .

McKendree, J., Stenning, K., Mayes, T., Lee, J., & Cox, R. (1998). Why observing a dialogue may benefit learning. Journal of Computer Assisted Learning, 14, 110-119.

Microsoft Agent. (June 11, 1998). Microsoft Corporation.
<http://www.microsoft.com/msagent/>.

Moreno, R., & Mayer, R. E. (1999). Cognitive principles of multimedia design: The role of modality and contiguity. Journal of Educational Psychology, 91, 358-368.

Murray, J. D. (1995). Logical connectives and local coherence. In R. F. Lorch & J. D. O'Brien (Eds.), Sources of coherence in reading (pp. 107-125). Mahwah, NJ: Erlbaum.

Person, N. K., Graesser, A. C., Kreuz, R. J., Pomeroy, V., & TRG, (2001). Simulating human tutor dialogue moves in AutoTutor. International Journal of Artificial Intelligence in Education, 12, 23-39.

Rosenthal, R. L., & Zimmerman, B. J. (1978). Social learning and cognition. New York: Academic Press.

Schober, M. F., & Clark, H. H. (1989). Understanding by addressees and overhearers. Cognitive Psychology, 21, 211-232.

Sweller, J. (1999). Instructional design in technical areas. Camberwell, Victoria, Australia: ACER Press.

Sweller, J., & Cooper, G. (1985). The use of worked examples as substitute for problem solving in learning algebra. Cognition and Instruction, 2, 59-89.

This research was supported by grants from DoD Multidisciplinary University Research Initiative (MURI) program administered by the Office of Naval Research (N00014-00-1-0600) and the National Science Foundation (SBR 9720314 and REC 0106965). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of ONR or NSF.

Table 1. The Means and Standard Deviations of the Number of Relevant, Related, and Irrelevant Propositions Written by Students on Discourse Overheard in Dialog and Monologue-like Formats.

Propositions	Dialog		Monologue	
	Mean	SD	Mean	SD
Irrelevant Propositions	2.48	3.93	2.36	3.00
Related Propositions	2.25	5.48	2.23	4.67
Relevant Propositions	15.41	10.44	8.82	8.09

Table 2. The Means and Standard Deviations of the Number of Relevant, Related, and Irrelevant Propositions for each Discourse Type and Dialog Condition

Dialog Discourse						
Dialog Condition	Relevant Propositions		Related Propositions		Irrelevant Propositions	
	Mean	SD	Mean	SD	Mean	SD
Deep Question	17.33	10.73	5.96	10.09	3.42	3.93
Shallow Question	14.88	8.55	4.21	6.93	2.63	5.11
Comment	13.63	7.61	6.63	9.15	3.21	3.89
Enhanced Monologue	14.38	10.66	3.58	4.86	2.63	3.90
Monologue-Like Discourse						
Dialog Condition	Relevant Propositions		Related Propositions		Irrelevant Propositions	
	Mean	SD	Mean	SD	Mean	SD
Deep Question	11.37	7.99	3.42	5.37	2.38	2.95
Shallow Question	13.13	9.38	7.17	10.86	3.42	3.98
Comment	12.08	9.58	4.04	6.44	3.42	4.50
Enhanced Monologue	17.08	9.57	2.17	5.20	2.50	4.62

Figure Captions

Figure 1. A screen shot on a topic concerned with computer speed.

The slide illustrates the difference between a 16-bit bus and a 32-bit bus. It features two diagrams of a computer system architecture, each with a RAM module at the top. The top diagram shows a 16-bit bus system with components labeled 'ARITHMETIC LOGIC', 'CONTROL', 'DECODE', 'BUS 16', and 'PREFETCH'. The bottom diagram shows a 32-bit bus system with components labeled 'ARITHMETIC LOGIC', 'CONTROL', 'DECODE', 'BUS 32', and 'PREFETCH'. A diagonal line separates the two diagrams. On the left, a cartoon man in a blue shirt stands behind a podium. On the right, a cartoon wizard in a blue robe with yellow stars and a blue hat stands. At the bottom, there is a navigation bar with four buttons: a left arrow, a right arrow, a question mark, and a right arrow.

Appendix A

Example of a Recall Question

Please write an explanation of how the various parts of the central processing unit interact with each other and with random access memory when an application program is running. Pretend you are writing to someone who does not know much about computers, or how they work. You may draw on anything you know about computers that will help you clarify your explanation.

Appendix B

Example Information Delivery and Discourse for Each Condition

Topic: The CPU and computer speed

Virtual tutor information delivery: There are many important issues to be considered when buying a computer. Compatibility is one important issue, but such issues as speed also need to be considered.

Virtual tutee monologue question condition: What components of computers can be modified by manufactures in order to increase speed and why is speed important?

Virtual tutee dialog deep question condition: Why is speed important?

Virtual tutee dialog shallow question condition: Is speed important?

Virtual tutee dialog comment condition: Speed is important.

Virtual tutor all conditions: The faster the CPU, the faster it can process data, exchange information with RAM, and communicate with peripherals.

Virtual tutee dialog deep question condition: How can a manufacturer increase the speed of the computer, what can they do to make it faster?

Virtual tutee dialog shallow question condition: Can you increase the speed?

Virtual tutee dialog comment condition: You can increase the speed.

Virtual tutor all conditions: Well, one thing manufacturers do is increase the clock speed of the computer.

Virtual tutee dialog deep question condition: But with a clock you want to know how accurate it is not how to make it go faster. Why would they want a clock to go faster, rather than increase its accuracy?

Virtual tutee dialog shallow question condition: Is clock speed important?

Virtual tutee dialog comment condition: Clock speed is important.