TESTS OF SOME MECHANISMS THAT TRIGGER QUESTIONS

Arthur C. Graesser, Cathy L. McMahen, and Brenda K. Johnson

Department of Psychology
Memphis State University
Memphis, TN 39152
bitnet: graesserac@memstvx1

Abstract

We have identified mechanisms that generate questions when individuals solve problems, comprehend text, and engage in conversations. Some of these mechanisms have been discussed in previous research in cognitive science and discourse processing, whereas other mechanisms were discovered when we analyzed videotapes of student-tutor interactions. The present study tested whether anomalous information causes an increase in questions when individuals solve mathematics problems and comprehend stories. College students were instructed to generate questions while they were solving problems (i.e., algebra and statistics) or while they were comprehending stories (e.g., fables and parables). There were several different versions of each problem or story: (1) complete original, (2) deletion of critical information, (3) addition of contradictory information, and (4) addition of irrelevant information. The deletion versions elicited the most questions whereas the original versions elicited the fewest questions; the addition versions were in-between. The validity of some of the question generation mechanisms is supported by the fact that these transformations of content caused an increase in questions.

Question generation has had a controversial status in cognitive science and education. At one extreme, there are models of cognition which postulate that question generation plays a critical role in comprehension, curiosity, and learning (Collins, 1988; Collins, Brown, & Larkin, 1980; Olson, Duffy, & Mack, 1985; Palinscar & Brown, 1984; Ram, 1990; Schank, 1986; Sternberg, 1987). According to Schank's SWALE model, for example, learning occurs when an individual observes an anomalous event and generates questions that lead to an explanation of the event (Schank, 1986).

This research was funded by grants awarded to the first author by the Office of Naval Research (N00014-88-K-0110 and N00014-90-J-1492).

A critical part of learning consists of generating illuminating questions and of transforming questions to become more illuminating. Collins (1988) has discussed how learning may be enhanced by having tutors ask students questions that expose the students' misconceptions during the course of answering the questions. The learning environment should also encourage students to ask questions and to model effective question asking (hopefully displayed by teachers). In the best of worlds, the learner would be an active, inquisitive, creative individual who energetically asks questions and insists on good answers.

The other end of the continuum provides a more pessimistic picture. It is well documented that student-generated questions are both infrequent and unsophisticated (Dillon, 1988; Flammer, 1981; Kerry, 1987). On the average, only one question is asked by a student per hour in a classroom setting. These questions are normally shallow questions that address the content and interpretation of explicit material rather than high-level questions that involve inferences, application, synthesis, and evaluation. A learner needs to master a significant amount of the material before the learner formulates good questions (Miyake & Norman, 1979). There are costs to posing questions in a classroom setting. The student is humiliated for revealing his or her lack of knowledge. The student imposes on the teacher, who rarely wants to be bothered by students' questions. Teachers frequently have trouble understanding the students' questions when the student has low domain-specific knowledge (Coombs & Alty, 1980), so the teacher ends up dismissing the question or answering the wrong question.

Unfortunately, teachers are normally poor role models for generating good questions. Less than 4% of the teacher-generated questions are higher-level demanding questions, i.e., those that involve inferences, the application of an idea to a new domain of knowledge, or the synthesis of a new idea from multiple information sources (Kerry, 1987). Teachers are rarely sophisticated Socratic tutors who ask carefully planned sequences of thought-provoking questions, thereby exposing the students' misconceptions and contradictions.

The nature of question generation mechanisms in humans has profound implications for expert systems, decision support systems, and intelligent tutoring systems (Buchanan & Shortliffe, 1984; Sleeman & Brown, 1982; Wenger, 1987). The users of these sophisticated computer systems may need to radically alter their styles of learning and thinking to the extent that the users are expected to be active, creative question askers. An inept question asking or question answering facility on a computer will rapidly kill any incentive for a user to ask the computer questions.

Given the above theoretical and practical issues surrounding question generation, cognitive scientists need to understand the mechanisms that trigger questions when individuals genuinely seek information. To the extent that we understand these question generation mechanisms, we can design programs that tap into these mechanisms and maximize the generation of good questions. Learning, creativity, and problem solving will be facilitated to the extent that good questions are asked by students and computer users.

Mechanisms that Generate Questions

Graesser, Person, and Huber (in press) identified four groups of mechanisms that generate questions. Some of these mechanisms exist in computer models of question generation, problem solving, and reasoning, such as AQUA (Ram, 1990; Reisbeck, 1988), SWALE (Schank, 1986), SOAR (Laird, Rosenbloom, & Newell, 1987), and SDDS (Klahr & Dunbar, 1988). Other mechanisms exist in theories of conversation (Allen, 1983; Clark & Schaefer, 1989), education (Collins, 1988; Sleeman & Brown, 1982), and question generation per se (Graesser, Lang, & Horgan, 1988; van der Meij, 1987). In addition to these theoretical contributions, Graesser et al. (in press) discovered question generation mechanisms when they analyzed transcripts of tutoring sessions on a variety of topics (e.g., research methods, mathematics, computer systems, climate, agriculture). The four groups of question generation mechanisms are summarized below.

(1) Correction of knowledge deficits. When individuals discover that their knowledge base is either incomplete or in error, they attempt to correct such knowledge deficits by asking questions from information sources that/who have useful, relevant information. Five question generation mechanisms are associated with this group:

(a) An obstacle in planning or problem solving. The question elicits information that removes or circumvents the obstacle.
(b) A decision among alternatives that are equally attractive. The question elicits information that breaks the tie.
(c) Glitch in an explanation of an event. The question elicits information that explains why an anomalous event occurred in a situation.
(d) Gap in knowledge that is needed for comprehension. The question elicits information that fills a gap in the questioner's knowledge base when he or she attempts to comprehend a written message, event, or conversation.
(e) Contradiction. The question elicits information that resolves the contradiction.

We were particularly interested in this first group of question generation mechanisms in the present study.

(2) Monitoring common ground. Successful communication requires that the speech participants keep track of what information is in the "common ground", i.e., knowledge that the speech participants believe they share. Speech participants frequently ask questions in order to update, verify, and monitor the knowledge in the common ground. Graesser et al. (in press) identified several question generation mechanisms which perform these functions: estimating or establishing common ground (e.g., "Do you know about X?"), confirming whether some belief is true ("Is X really true?"), and gauging whether the questioner or answerer comprehends something ("Do you understand what I'm saying?", "Am I understanding what you are saying?").

(3) Social coordination of action. Questions are frequently generated to coordinate actions of speech participants, to manipulate people, or to get permission to do something. Graesser et al. identified the following mechanisms in this group: indirect requests (e.g., "Would you do X?"), indirect advice ("Why don't you do X?"), asking permission ("Can I do X?"), offers ("Would you like me to do X for you?"), and negotiations ("If I do X, would you do Y?").

(4) Control of conversation and attention. Questions are frequently asked to monitor the flow of conversation among speech participants. This includes getting a speaker's attention, maintaining attention on a speaker, and changing speakers. Graesser at al. identified the following mechanisms in this group: greetings, replies to summons, changing speakers, maintaining focus on an agent's actions, rhetorical questions, and gripes.

The purpose of the present study was to test some of the question generation mechanisms in the first group, i.e., correction of knowledge deficits. In two experiments, college students were instructed to generate questions while they solved mathematics problems or while they comprehended stories. There were alternative versions of each problem or story. One version was the original complete problem or story. The other versions were transformations of the original version that were designed to elicit additional questions. A critical piece of information was removed in deletion versions. Addition versions contained an extra piece of information that introduced either a contradiction or an irrelevancy. If the question generation mechanisms are valid, then the transformed versions should yield more questions than the original versions. Moreover, a subset of the questions should address the particular transformations.

Question Generation in the Context of Mathematics Problems.

Methods. Twelve original mathematics problems were selected from textbooks. Six of the problems were word problems from a basic algebra text at the seventh grade level; the other six were statistics problems from a college-level statistics textbook. In addition to the original version of each problem, we prepared a deletion version and an addition version that contained irrelevant information. An example of these three versions is presented below.

ORIGINAL. A bottle of Harvest Time Apple Juice contains 64 ounces and costs 99 cents. Farm Fresh Juice is available in bottles that contain 1 gallon for $1.88 each. Which is the better buy?

DELETION. A bottle of Harvest Time Apple Juice costs 99 cents. Farm Fresh Juice is available in bottles that contain 1 gallon for $1.88 each. Which is the better buy?

ADDITION. A bottle of Harvest Time Apple Juice contains 64 ounces and costs 99 cents. There are coupons which give the customer a 10% savings on all items. Farm Fresh Juice is available in bottles that contain 1 gal for $1.88 each. Which is the better buy?

The subjects were 28 college students who were enrolled in a Research Methods course at Memphis State University. Each subject received 4 problems in each of the 3 versions, yielding 12 problems altogether. The assignment of problem versions to subjects was counterbalanced so that approximately the same number of subjects received each version of a particular problem. Each subject received a 12-page booklet, with one problem per page.

The subjects were instructed to work on each problem in three phases, which were timed by the experimenter. The subjects read the problem during phase 1 for 30 seconds. The subjects wrote down questions that came to mind about the problem during phase 2 (90 seconds). The subjects solved the problem and generated additional questions during phase 3 (150 seconds). The subjects were given 30 seconds of rest before starting on the subsequent problem. Therefore, 5 minutes was allocated to each of the 12 problems.

Results. We scored the number of questions that each subject generated for each problem. An expression counted as a question if it was in an interrogative form, as opposed to an assertion or other type of speech act. Questions can be identified by trained judges with a high degree of reliability (Graesser et al., in press). The mean number of questions per problem significantly differed among versions, with means of 1.81, 2.49, and 2.09 in the original, deletion, and addition versions, respectively, $F(2, 54)= 5.61$, $p < .05$. Planned comparisons revealed the following differences: deletion > addition = original. Therefore, the deletion transformations had a more robust impact on question generation than did the addition transformations.

In a follow up analysis, we scored those questions that addressed the transformations in the deletion and addition versions. For example, a question that would address the deletion version would be "How big is the bottle of the Harvest Time apple juice?"; a question that would address the addition version would be "Why is there this information about coupons when it makes no difference as to the true value?" The mean number of transformation-relevant questions per problem was significantly higher in the deletion versions than the addition versions, .81 versus .41, respectively, $t(27)= 4.32, p < .05$. Both the .81 and the .41 significantly differed from 0 (i.e., the baseline level of the original version). Therefore, both types of transformations caused a significant increase in the number of transformation-relevant questions.

Question Generation in the Context of Stories.

Methods. The original stories were 5 parables and 5 fables selected from anthologies of classical or famous writers. The stories had a short length of approximately 100 words. An example story is presented below.

> A hunter caught a wild rooster and brought it to the city for sale. A merchant asked: "What kind of bird is this? " "A phoenix," cheated the hunter. "That is a lucky symbol, which I see for the first time. Do you want to sell it to me?" "Sure, if you pay a high price." The merchant gave him 20 tael because he wanted to bring it to the king. But on the way the rooster died. He cried loudly, because he could not bring it to the king. People heard of this. Later, even the king knew of it. The king then gave him 200 gold pieces for his kindness, although he did not know it had only been a rooster.

There were five versions of each story as specified below.

ORIGINAL. This was the original, complete story.

DELETION. A critical piece of information was deleted; the point or moral of the story could not be grasped without this information. In the context of the example story, we deleted this sentence: "But on the way the rooster died."

ADDITION-CONTRADICTION. Information was inserted that contradicted a major idea in the original story. For example, we added this sentence: "The merchant had the bird shipped to his daughter."

ADDITION-SUBTLE-IRRELEVANCY. The additional information was an episode that could occur in the story, but was not relevant to the main plot and point of the story. For example, the story in this condition reads: "The merchant locked all the doors properly. But on the way the rooster died."

ADDITION-SALIENT-IRRELEVANCY. The additional information was an episode that would probably not occur in the story, and was not relevant to the main plot and point of the story. For example, the following statement was added to the text: "The teacher showed him how to do the equations."

The subjects were 24 students enrolled in an Introductory Psychology class at Memphis State University. Each subject completed a 10-page booklet, with one page assigned to each story. Each subject had one fable assigned to each of the 5 versions and one parable assigned to each of the 5 versions. The materials were counterbalanced so that each story had approximately the same number of subjects assigned to each of the 5 versions. The subjects' task was to read each story and to generate questions that came to mind while reading the stories.

Results. The mean number of questions per story was 3.13, 3.63, 3.21, 3.65, and 3.54 in the original, deletion, addition-contradiction, addition-subtle-irrelevancy, and addition-salient-irrelevancy conditions, respectively. The mean of the four transformation conditions was significantly higher than the mean of the original condition, $F(1, 23)= 4.67, p < .05$; the four transformation conditions did not significantly differ from one another.

We computed the mean number of questions that were relevant to the transformations of the story content. The four transformation conditions significantly differed on this measure, with means of .77, .47, .38, and .67 in the deletion, addition-contradiction, addition-subtle-irrelevancy, and addition-salient-irrelevancy conditions, respectively. All of these scores significantly differed from a zero baseline (which would correspond to the original version that had no transformations). Therefore, the transformations in content caused an increase in generated questions and a subset of the questions addressed the transformations. Our data suggest that the deletions and additions of salient irrelevancies were more obvious to the students than were the contradictions and additions of subtle irrelevancies.

Conclusions

We found that the transformed versions of the mathematics problems and stories elicited more questions than did the original versions. Moreover, in the transformed versions, a subset of the questions addressed the transformations. These findings are consistent with the models of question generation that were discussed earlier. For example, Schank's SWALE model predicts that anomalous events trigger questions (Schank, 1986). The contradictions and irrelevant statements that were added in our transformed versions would be construed as anomalous information. Our findings support the claim that some of the question generation mechanisms identified by Graesser at al. (in press) are causally related to the volume and qualitative characteristics of generated questions. In particular, support was found for those question generation mechanisms that involve "corrections of knowledge deficits."

Graesser at al. (in press) identified five question generation mechanisms associated with corrections of knowledge deficits. Four of these five mechanisms were manipulated by the transformations in this study. The deletion transformations involved both (a) an obstacle in a plan or problem solving and (b) a gap in knowledge that is needed for comprehension. Clearly, the students faced an insurmountable obstacle when a problem omitted information that is critical for solving the problem (or alternatively, comprehending the point of a story). The transformation that added a contradiction obviously involved Graesser et al.'s contradiction mechanism. The transformations that added irrelevancies involved either (a) a gap in knowledge that is needed for comprehension or (b) a glitch in an explanation of an event. There is a pragmatic assumption that all information is relevant and important when short stories and problems are constructed. Students presumably would not believe that they completely comprehended a story or problem unless they had some explanation of an irrelevancy.

According to the results of the present study, the above question generation mechanisms are causally related to the volume and qualitative questions that students generate. Person (1990) has also reported that these question generation mechanisms naturally occur when students generate questions in a tutoring session on research methods (e.g., statistics, measuring variables, testing hypotheses, factorial designs). Nearly half of the student-generated questions in Person's videotaped tutoring sessions were in the major category of "corrections in knowledge deficits." Therefore, there is support for these question generation mechanisms both in controlled laboratory environments and in naturalistic environments.

References

Allen, J. 1983. Recognizing Intentions from Natural Language Utterances. In M. Brady, and R.C. Berwick (Eds.). *Computational Models of Discourse.* Cambridge, MA: MIT Press.

Buchanan, B.G., and Shortliffe, E.H. eds. 1984. *Rule-Based Expert Systems.* Menlo Park, CA: Addison-Wesley.

Clark, H.H., and Schaefer, E.F. 1989. Contributing to Discourse. *Cognitive Science* 13:259-294.

Collins, A. 1988. Different Goals of Inquiry Teaching. *Questioning Exchange* 2: 39-46.

Collins, A., Brown, J.S., and Larkin, K.M.. 1980. Inference in Text Understanding. In R.J. Spiro, B.C. Bruce, and W.F. Brewer (Eds.). *Theoretical Issues in Reading Comprehension* (pp.385-407). Hillsdale NJ: Erlbaum.

Coombs, M.H., and Alty, J.L. `1980. Face-to-Face Guidance of University Computer Users-II. Characterizing Advisory Interactions. *International Journal of Man-Machine Studies* 12:407-429.

Dillon, J.T. 1988. *Questioning and Teaching: A Manual of Practice.* New York: Teachers College Press.

Flammer, A. 1981. Towards a Theory of Question Asking. *Psychological Research* 43: 407-420.

Graesser, A.C., Lang, K.L., and Horgan, D. 1988. A Taxonomy for Question Generation. *Questioning Exchange* 2(1): 3-15.

Graesser, A.C., Person N.K., and Huber J.D. (in press). Mechanisms That Generate Questions. In T. Lauer, E. Peacock, and A.C. Graesser (Eds.). *Questions and Information Systems.* Hillsdale, HJ: Erlbaum.

Kerry, T. 1987. Classroom Questions in England. *Questioning Exchange* 1(1):32-33.

Klahr, D and Dunbar, K. 1988. Dual Search Space During Scientific Reasoning. *Cognitive Science* 12(1):1-48.

Laird, J., Rosenbloom, P. S., and Newell, A. 1987. SOAR: An Architecture for General Intelligence.

Artificial Intelligence 33(1): 1-64.

Miyaki, N. and Norman, D.A. 1979. To Ask a Question, One Must Know Enough to Know What is Not Known. *Journal of Verbal Learning and Verbal Behavior* 18:357-364.

Olson, G.M., Duffy, S.A., and Mack, R.L. 1985. Question Asking as a Component of Text Comprehension. In A.C. Graesser and J.B. Black (Eds.). *The Psychology of Questions*. (pp. 219-226). Hillsdale NJ: Erlbaum.

Palinscar, A.S., and Brown, A.L. 1984. Reciprocal Teaching of Comprehension-Fostering and Comprehension-Monitoring Activities. *Cognition and Instruction* 1(2):117-175.

Person, N.K. 1990. The Documentation of Questioning Mechanisms and Types of Questions in Tutoring Protocols. Unpublished master's thesis, Memphis State University, Memphis, TN.

Ram, Ashwin. 1990. Decision Models: A Theory of Volitional Explanation. In Proceedings of the Twelfth Annual Conference of the Cognitive Science Society, 198-205. Hillsdale NJ: Erlbaum.

Reisbeck, C.K. 1988. Are Questions Just Function Calls? *Questioning Exchange* 2 (1):17-24.

Schank, R.C. 1986. *Explaining Problems: Understanding Mechanically and Creatively*. Hillsdale, NJ: Lawrence Earlbaum Associates.

Sleeman, D.H. and Brown, J.S. eds. 1982. *Intelligent Tutoring Systems*. New York: Academic Press.

Sternberg, R.J. 1987. Questioning and Intelligence. *Questioning and Exchange* 1:11-13.

van der Meij, H. 1987. Assumptions of Information-Seeking Questions. *Questioning Exchange* 1(2):111-117.

Wenger, E. 1987. *Artificial Intelligence and Tutoring Systems*. Los Altos, CA: Morgan Kaufman.