Individual Differences in Reasoning about Broken Devices: An Eye Tracking Study

Shulan Lu (shulanlu@hotmail.com)
Department of Psychology, University of Memphis, 3693 Norriswood
Memphis, TN 38152 USA

Brent Olde (baolde@memphis.edu)
Department of Psychology, University of Memphis, 3693 Norriswood
Memphis, TN 38152 USA

Elisa Cooper (culdihavben@hotmail.com)
Department of Psychology, University of Memphis, 3693 Norriswood
Memphis, TN 38152 USA

Arthur C. Graesser (a-graesser@memphis.edu)
Department of Psychology, University of Memphis, 3693 Norriswood
Memphis, TN 38152 USA

Abstract

We examined the questions college students ask when everyday devices malfunction. Our investigation of the explanatory reasoning processes is organized around the central theme of question asking. PREG, a model of human question asking, predicts when and what types of questions are asked by humans while comprehending expository texts. PREG has two salient predictions. First, deep comprehenders should ask questions that converge on plausible faults. Second, eye movements should converge on those likely faults. An eye tracking study supported these predictions. The present research supports the claim that question asking and eye tracking are two excellent indicators of device comprehension in the context of breakdown scenarios.

Introduction

Deep comprehension of everyday devices can be manifested in a number of tasks and measures. For example, most colleagues would agree that the deep comprehenders construct coherent representations of functioning devices, draw appropriate inferences, answers explanation-based questions correctly (e.g., why, how, what-if, what-if-not), and solve transfer problems that apply their understanding (Gentner & Stevens, 1983; Graesser, Singer, & Trabasso, 1994). This study investigates the manifestations of deep comprehension of devices that are not as widely recognized. We believe that deep comprehension is required when devices breakdown, when eyes fixate on likely faults, and when questions are asked about likely faults. (Graesser, Olde, & Lu, in press; Graesser, Olde, Pomeroy, Whitten, Lu, & Craig, in press).

Question asking and its role in understanding texts and stories is well-documented (Graesser & McMahen, 1993; Kass, 1992; Schank, 1986, 1999; Ram, 1994). The literature has consistently suggested that the understanding of a story is achieved by identifying the questions raised by the story and then searching for the answers in the story. Question asking has many potential functions in reasoning and problem solving. For instance, question asking is often affiliated with searches, comparisons, explanations, predictions and several other cognitive processes. It is reasonable to say that comprehenders would have more questions as they reason through an expository text and that their questions would manifest comprehension depth. For example, it is very hard to imagine that a shallow comprehender could ask questions addressing the critical causal components of an event. In this article, we examine a model of question asking in the context of understanding technical texts. More specifically, we are interested in the model's predictions about deep versus shallow comprehension of everyday devices as revealed by question asking. It is suggested that question asking opens a window for viewing the sub-processes involved in understanding (Ram, 1994), such as the retrieval of explanations in long term memory and the search of information from a display. However, there has been very little empirical research that documents the relationships among comprehension, question asking, and information search (as reflected in eye tracking). Therefore, we conducted a study in which eye tracking was measured while college students generated questions when confronted with a breakdown scenario.

PREG: A Model of Question Asking

Several computational models of question asking have been constructed in the context of story understanding. These models were capable of generating questions with respect to the goals and sub-goals of a story. However, an adequate model of question asking should be capable of predicting when and what types of questions humans ask as they comprehend expository texts as well as other types of learning material. Otero and Graesser (in press) developed a PREG model of question asking that attempts to capture these question asking mechanisms in detail. The general assumption of the model is that clashes between text input and a reader's existing world knowledge trigger question generation (Graesser, Olde, Pomeroy, et al., in press; Otero et al., in press). Questions are constructed when readers come across information in a text that presents contradictions, anomalies, obstacles to goals, discrepancies, constrasts and other triggers of potential cognitive disequilibrium (Graesser et al., 1993; Graesser & Person, 1994).

The discrepancies between input and world knowledge can be associated with the different levels of representations, ranging from shallow to deep (Britton & Graesser, 1996; Gentner et al., 1983; Graesser, Millis, & Zwaan, 1997; Kieras & Bovair, 1984; Kintsch, 1998). The surface code, which is at the shallowest level, keeps the wording and syntax of a text in a verbatim form. As for the visual modality, it preserves the low-level lines, angles, sizes, shapes, and textures of the picture. The textbase, which is at the intermediate level, in essence is a propositional representation that maintains the meaning of the explicit text and the pictures. The mental model, which is at the deepest level, captures the referential content of the text. When applied to everyday devices, this would include:

1. the components of the electronic or mechanical system;
2. the spatial arrangement of components;
3. the causal chain of events when the system successfully unfolds;
4. the mechanisms that explain each causal step;
5. the functions of the device and device components;
6. the plans of agents who manipulate the system for various purposes.

Quite clearly, a rich set of knowledge structures needs to be constructed when an adult comprehends a device at a deep level.

According to the PREG model, conceptual graph structures are adopted to encode a chronology of events and states that happen during the course of device motion. The conceptual graph structure is not built arbitrarily, but is comprised of a set of categorized nodes which denote concepts and proposition-like descriptions in the text and corresponding visual-spatial information. These nodes are connected by arc categories such as ENABLE, CAUSE, PROPERTY, REASON and OUTCOME. In addition, most arcs are directed with a source node and an end node.

It is assumed that pictorial and textual information is incorporated in a single underlying representation. Empirical studies show that most readers are capable of alternating between picture and text and that the text dominates the reading process when illustrated texts are comprehended (Bagget & Graesser, 1995; Hegarty & Just 1993).

Individual Differences in Question Asking

The current study examines the questions that college students ask when an everyday device malfunctions. For example, consider a cylinder lock and the following breakdown scenario: the key turns, but the bolt does not move. According to PREG, understanding is manifested when a device breaks, not when it is running smoothly. Thus PREG predicts that deep comprehenders should ask good questions that converge on likely faults. More specifically, these questions should tap the nodes in the conceptual graph structure that are the plausible causes of the malfunction. To test this hypothesis, Graesser, Olde, Pomeroy et al. (in press) conducted a study in which 108 participants first read an illustrated text, then were provided a breakdown scenario, and then generated questions. After completing the question asking task, an objective comprehension test on the devices were administered. A battery of tests that measure cognitive abilities and personality were administered in the end.

The results confirmed the hypothesis: Good comprehenders generate high quality questions that focus on plausible faults of the breakdown. Follow-up multiple regression analysis further suggested that ASVAB (the Armed Services Vocational Aptitude Battery, Department of Defense, 1983) technical score was the primary predictor of both deep comprehension and question quality.

Given that technical scientific knowledge turned out to be a robust predictor of device comprehension, we conducted a qualitative analysis of the questions asked by participants with high (upper 33% of distribution) versus low technical knowledge (lower 33% of the distribution). The questions generated by participants with high technical scores tend to converge on the fault components and address the causal connections between parts, processes, and relations that are in the chain of breakdown. The questions asked by low technical participants tend to be diffuse. That is, most of the components in a system were addressed in the hope that it might turn out to be pertinent instead of converging on 1 or 2 parts. Their questions rarely were

elaborations on the causal links addressing the malfunction.

Since the above patterns emerged, we were curious to know whether there were systematic differences in the eye movement patterns between individuals with different levels of cognitive abilities. That is the focus of the present study.

Question Asking and Control of the Eye

Eye movements provide an important window for understanding the cognitive processes and representations that play a role in a particular cognitive task. However, no one has investigated the relationship between eye movements and the cognitive components in question asking. PREG predicts that eye movements should converge on the likely faults. As far as individual differences in the eye movement are concerned, the following hypotheses could be directly generated from the PREG model. First, deep comprehenders are expected to have a high density of eye fixations occur at words, objects, parts, and processes that are at the source of cognitive disequilibrium (e.g., anomalies, contradictions, broken parts, contrasts, missing components, and so on), while shallow comprehenders should indiscriminately scan the regions of the illustrated text. A sufficient amount of technical knowledge is necessary for identifying anomalies in a system. Thus, technical knowledge and other indices of deep comprehension should be positively correlated with measures of the fixations that assess the extent to which a comprehender focuses on fault areas.

Method

Participants
The participants were forty college students at the University of Memphis. The students participated for course credit in an introductory psychology class.

Illustrated Texts and Question Asking Tasks
The participants read 5 illustrated texts on everyday devices: a cylinder lock, an electronic bell, a car temperature gauge, a toaster, and a dishwasher. The illustrated texts were extracted from Macaulay's (1988) book, *The Way Things Work.* These were the same devices that were used in the Graesser, Olde, Pomeroy et al., (in press) study, except that the clutch was dropped from the current study; it was extremely difficult for participants to differentiate and label the individual teeth in the wheels of the clutch mechanism.

As in the Graesser, Olde, Pomeroy et al., (in press) study, there were five trials, each of which consisted of two phases. The participant first read an illustrated text for 3 minutes, which was displayed on a computer monitor. After the reading phase, the breakdown description was presented either above or to the left of the illustrated text and the participants began the question asking phase (while the illustrated text remained on the screen). The participants asked questions aloud for 90 seconds during this phase and the protocol was recorded. The previous study had participants generate questions in writing whereas the present study collected spoken questions. Each participant furnished question asking protocols for all 5 devices. The assignment of devices to test order was counterbalanced across the 40 participants with a Latin square.

Device Comprehension Test
The participants subsequently completed a 30-question test of device comprehension (5 devices x 6 three-alternative, forced-choice question per device). All 30 questions were generated from a theoretical framework in qualitative physics (Forbus, 1984). Suppose there are N components in a system and their states are delineated as either inhibitory, excitatory, or neutral as affected by other components in the causal network. The test questions are concerned with how tweaking one component A has an impact on another component B. An example question constructed according to the three possible states is as follows:

What happens to the pins when the key is turned to unlock the door?
   (a) they rise
   (b) they drop
   (c) they remain stationary (correct answer)

It is not likely that participants will be able to answer such questions correctly without deep comprehension of the devices. It is reasonable to predict that deep comprehenders will be able to trace the causal antecedents and causal consequences of the events (Graesser & Bertus, 1998) and shallow comprehenders will lose track of causal connections. The device comprehension test was thus designated as the gold standard for deep comprehension. It is predicted that performance on the device comprehension test should positively correlate with the quality of the questions that get asked and also with fixations on the faults of a breakdown scenario. The device comprehension score could vary from 0 to 30. A score of 10 would be chance performance if there were no sophisticated guessing or background knowledge.

Battery of tests of individual differences
The participants completed assessments of the following measures of individual differences: four scales of technical knowledge (mechanical comprehension, electronics, general science, auto & shop) extracted from ASVAB (Department of Defense, 1983), and additional tests of spatial reasoning (Bennet,

Seashore, & Wesman, 1972) and openness (Costa & McCrae, 1991). These were included, as they were the statistically significant predictors of deep comprehension and question asking in the Graesser, Olde, and Pomeroy et al. (in press) study.

Recording of eye tracking and question asking
Eye movements were recorded by a Model 501 Applied Science Laboratory eye tracker. There was a magnetic head tracker so the participants could move the head during data collection. The participants were calibrated before they started the experimental session of reading the illustrated texts and asking questions. During calibration, the participants viewed 9 points on the computer display and a computer recorded the x-y coordinates. The calibration process took 10-15 minutes. Participants were dismissed if they wore glasses, but the equipment could accommodate contact lenses.

The experimental session was videotaped and audio recorded. The VCR camera focused on a scene monitor screen, on which the illustrated text being viewed by the participant and the trace of the participant's eye movements were mirrored. The VCR recorded both the illustrated text and a superimposed image of what the left eye was focusing on. The superimposed image showed the locus of (a) the focus of the eye and (b) an X-Y axis with the 0-0 point at the center of the focus. The voice of the participant was recorded on the VCR so that the spoken questions could be transcribed. This set-up allowed us to record and review (a) the contents of the computer display, (b) the focus of the left eye, and (c) the voice of the student asking questions.

Computer software was available to record eye movements at a fine-grained level. The software produces area plots for specific areas of interest. In particular, we were interested in the portions of eye fixations focusing on the areas of interest associated with faults. These faults were sometimes in the text and sometimes in the picture.

The following measures were scored on the think aloud protocols collected in the question asking task.

Volume of questions: The number of questions that were asked in the question asking task.
Question Quality: The number or proportion of questions that referred to a plausible explanation of the breakdown.

Trained judges coded the verbal protocols with an acceptable level of reliability.

Results and Discussion
Descriptive Statistics
Table 1 presents means and standard deviations for the measures collected in this study. The ASVAB measures were comparable to normal college student populations. Scores of spatial reasoning were not significantly different from the scores for college students reported in Bennet et al. study (1972).

Table 1: Descriptive Statistics on Measures of Individual Differences

Measures	Mean	SD
Mechanical	13.9	6.0
Electronics	9.3	4.3
Auto & Shop	10.0	5.3
General Science	16.9	4.7
Spatial	23.8	15.2
Openness	52.1	9.2
Gender	1.25	.44

The Coordination between Question Asking and Eye Movements
It is important to examine the device comprehension scores first. The mean device comprehension score was 18.6 (SD = 4.6), which is 62% of the questions being answered correctly. As would be expected, device comprehension was significantly correlated with the number of fault questions asked (r = .45, p < .01). In addition, device comprehension showed a high positive correlation with all the ASVAB measures, in particular, ASVAB technical knowledge (r = .54) and general science (r = .60).

The data supported PREG's predictions. There were 29.5 fixations (SD = 11.1) on plausible faults per device, or 9.3 seconds (SD = 3.9) out of 90 seconds. 11.5% of the eye fixations and 10.4% of the time focused on the fault areas. The index of deep comprehension, i.e., device comprehension score, was significantly correlated with the eye tracking measure of number of fixations on faults (r = .43).

We examined the role of question asking in eye tracking at this point. The data showed strong evidence that there was consistent coordination between question asking and eye movements. Furthermore, there were significant differences in the measures of eye tracking between participants who had a relatively high number of fault questions and those who had a low number of questions, e.g., number of fixations on faults (167.6 with SD = 42.2 versus 127.5 with SD = 61.0, for high versus low).

Given the importance of cognitive abilities during device comprehension, we explored which measures of cognitive ability are more capable of discriminating deep comprehension versus shallow comprehension as measured by question asking and eye movements. General science turned out to be a significant predictor of the eye tracking measure and the number of fault questions asked. The participants who scored higher on general science had more questions on faults (7.7 with

SD = 2.7 versus 6.0 with SD = 3.0, for high versus low), and had bigger number of fault fixations (173.1 with SD = 48.7 versus 122.0 with SD = 50.9).

Openness is one of the "big five" personality factors: neuroticism, extroversion, openness, agreeableness, and conscientiousness (Costa et al., 1991). The subscale of openness attempts to capture creativity. A t-test on the eye tracking measure and question asking measure showed that there were significant differences between participants with high openness scores versus those with low openness scores: number of fixations on faults (167.8 with SD = 49.6 versus 127.3 with SD = 55.0, for high versus low), and number of fault questions (8.2 with SD = 2.9 versus 5.5 with SD = 2.4). The data suggested that openness was a robust predictor of question asking and subsequently affected the patterns of eye movements.
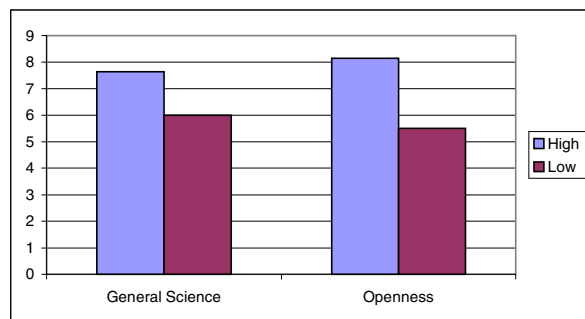

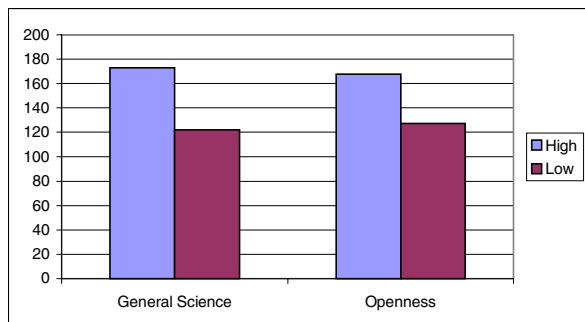Figure 1: The number of questions on faults by high versus low on general science and openness scales.

Figure 2: The number of fixations addressing faults by high versus low on general science and openness scales.

There have been two different views concerning the relationships between eye movements and ongoing cognitive processes (Rayner, Reichle, & Pollatsek, 1998). One is that eye movements are mainly driven by oculomotor processes; the other is that there is a correspondence between eye movements and cognitive processes. The data quite clearly suggested the coordination between question asking and eye movements. The question at this point was to what extent question asking was in control of eye movements. The ideal method of addressing this question would be structural equation modeling. However, given the number of participants in the present study, we could only conduct a partial correlation analysis.

If our PREG model is correct, there should not be a significant correlation between the number of questions addressing faults and readers' technical knowledge after partitioning out the variance of question asking. Subsequent data analysis supported the hypothesis that question asking is in control of eye movements rather than eye tracking guiding question asking. When the variable (number of fault questions) was controlled, the correlation between technical knowledge and the number of fixations on faults approached 0 (r = .14, p < .05). However, controlling the variance of eye movements did not affect the correlation between technical knowledge and the number of fault questions. They remained significantly correlated (r = .39, p < .05). The results are consistent with some recent findings which suggest that some cognitive processes are fast enough to affect eye movements (Rayner et al., 1998).

In short, it appears that individuals with high scores on general science and openness are most capable of asking questions about the anomalous information in a system. Subsequently these individuals move their eyes to the plausible fault areas and verify their reasoning about the breakdown scenario. On the other hand, individuals, who are low on the general science and openness scales, tend to be less sensitive to the contradictions when they arise. Thus they resort to the strategy of scanning all the regions of a text in the hope of hitting the target.

Conclusion

The current research has demonstrated the usefulness of the eye tracking data for studying the cognitive components in question asking. The analysis of eye fixations provides an account of how people with different levels of cognitive ability and different types of personality reason through the device malfunction and how their explanatory reasoning processes center around anomaly detection and question asking.

Acknowledgements

References

Baggett, W. B., & Graesser, A. C. (1995). Question answering in the context of illustrated expository text. *Proceedings of the 17th Annual Conference of the Cognitive Science Society* (pp. 334-339). Hillsdale, NJ: Lawrence Erlbaum.

Bennet, G. K., Seashore, H. G., & Wesman, A. G. (1972). *Differential aptitude test: Spatial relations., Form T.* New York: Psychological Corporation.

Britton, B., & Graesser, A. C. (1996) (Eds.). *Models of understanding text.* Hillsdale, NJ: Erlbaum.

Costa, P. T., & McCrae, R. R. (1991). *NEO: Five Factor Inventory.* Odessa, FL: Psychological Assessment Resources.

Department of Defense (1983). *Armed Services Vocational Aptitude Battery, Form 12a.* Washington, D.C.: Department of Defense.

Forbus, K. (1984). Qualitative process theory. *Artificial intelligence, 24*, 85-168.

Gentner, D., & Stevens, A. L. (1983) (Eds.). *Mental models.* Hillsdale, NJ: Erlbaum.

Graesser, A. C., Bertus, E. L. (1998). The construction of causal inferences while reading expository texts on science and technology. *Scientific Studies of Reading, 2*, 247-269.

Graesser, A. C., & McMahen, C. L. (1993). Anomalous information triggers questions when adults solve quantitative problems and comprehend stories. *Journal of Educational Psychology, 85*, 136-151.

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

Graesser, A. C., Singer, M., Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review, 101*, 371-395.

Graesser, A. C., Millis, K. K., Zwaan, R. A. (1997). Discourse comprehension. *Annual Review of Psychology, 48*, 163-189.

Graesser, A. C., Olde, B., & Lu, S. (in press). Question-driven explanatory reasoning about devices that malfunction. In T. Filjak (Ed.), *Proceedings of the 36th International Applied Military Psychology Symposium.*

Graesser, A. C., Olde, B., Pomeroy, V., Whitten, S., Lu, S., & Craig, S. (in press). Inferences and questions in science text comprehension. In J. Otero and M. Helena (Eds.), *Science text comprehension.*

Hegarty, M., & Just, M. A. (1993). Constructing mental models of machines from text and diagrams. *Journal of Memory and Language, 32*, 717-742.

Kass, A. (1992) Question-asking, artificial intelligence, and human creativity. In T. Lauer, E. Peacock, & A. C. Graesser, (Eds.), *Questions and information systems.* Hillside, NJ: Erlbaum.

Kieras, D. E., & Bovair, S. (1984). The role of a mental model in learning to operate a device. *Cognitive Science, 8*, 255-274.

Kintsch, W. (1998). *Comprehension: A paradigm for cognition.* New York, NY: Cambridge University Press.

Macaulay, D. (1988). *The way things work.* Boston, MA: Houghton Mifflin.

Otero, J., & Graesser, A. C. (in press). PREG: Elements of a model of question asking. *Cognition and Instruction.*

Ram, A. (1994) AQUA: Questions that drive the explanation process. In R.C. Shank, A. Kass, & C.K. Riesbeck (Eds.), *Inside Case-Based Explanation.* Hillsdale, NJ: Erlbaum.

Rayner, K, Reichle, E. D. & Pollatsek, A. (1998). Eye movement control in reading: An overview and model. In G. Underwood (Eds.), *Eye guidance in reading and scene perception.* Oxford, England: Elsevier.

Schank, R. C. (1986). *Explanation patterns: Understanding mechanically and creatively.* Hillsdale, NJ: Erlbaum.

Schank, R. C. (1999). *Dynamic memory revisited.* Cambridge: Cambridge University Press.