OFFICE OF NAVAL RESEARCH

END-OF-YEAR REPORT

PUBLICATIONS/PATENTS/PRESENTATIONS/HONORS/STUDENTS REPORT

for

GRANT: N00014-98-1-0331

PR Number 98PR04101-00

Question-driven Explanatory Reasoning About Devices that Malfunction

Principle Investigator: Arthur C. Graesser

Department of Psychology
The University of Memphis
CAMPUS BOX 526400
Memphis, TN 38152-6400
(901) 678-2742 (p) (901) 678-2579 (fax)
a-graesser@memphis.edu

Submitted July 14, 1999

Reproduction in whole, or in part, is permitted for any purpose of the United States Government.

This document has been approved for public release and sale, its distribution is unlimited.
PART I

a. Number of papers submitted to refereed journals, but not published: 2

b. Number of papers published in refereed journals: 12


c. +Number of books or chapters submitted, but not yet published: 4

d. +Number of books or chapters published: 20

Books:


Chapters:


**Published Conference Proceedings:**


e. **Number of printed technical reports/non-referred papers: 2**


f. **Number of patents filed: 0**

g. **Number of patents granted: 0**

h. **Number of invited presentations: 8**

Graesser, A.C. (July, 1998). Eight cool things from the collocated conferences. Invited panel at the Fifteenth National Conference on Artificial Intelligence, Madison, WI.


**i. +Number of submitted presentations: 12**


j. +Honors/Awards/Prizes for contract/grant employees: 3

Board of Visitors Eminent Faculty Award, University of Memphis. May, 1999.

Editor of the journal Discourse Processes, 1997-2004.

1999 Program Chair for Division C (Learning and Instruction) of the American Educational Research Association Conference, Montreal Canada, April 1999.

k. +Total number of Full-time equivalent Graduate Students and Post-Doctoral associates supported during this period, under the PR number:

Graduate students: 2 (1 female, 1 male, 0 Asian, 0 Minority)
Post-doctoral Associates: 0

l. Other funding

All of the following projects investigate question asking and question answering mechanisms, which is the direct focus of the ONR grant.

National Science Foundation, Simulating tutors with natural dialog and pedagogical strategies, 1997-2000, $900,000 (Dr. Graesser is PI, $210,000 received this year). This project involves developing an automated computer tutor (called AutoTutor) that asks and answers questions; question asking and answering is the direct focus of the ONR grant, so the projects are quite related. Also, both projects involve the technical domains; the NSF grant tutors students on computer literacy whereas the ONR grant investigates mechanical and electronic devices.
United States Bureau of Census. QUEST questionnaire evaluation tool. 1998-1999. $58,512 (Dr. Graesser is PI; $30,000 received this year). This project involves building a computer tool that critiques questions on surveys and questionnaires. Question asking and answering is the direct focus of the ONR grant so these projects are quite related.

National Institutes of Health. Smoking cessation in the elderly. 1998-2001. $1,600,000 (Dr. Bob Klesges is PI). As a co-PI, my role is to recommend changes on the question asking, question answering, and discourse interaction styles of health educators who try to get elderly smokers to stop smoking. Question asking and answering mechanisms are investigated in both the ONR and NIH projects.

Office of Naval Research. Intelligent distributed agents. 1998-2002. $1,500,000 (Dr. Stan Franklin is PI). As a co-PI, my role is to advise graduate students how to design question asking and question answering interfaces on software agents, and how to evaluate the system performance.
PART II

a. Principal Investigator: Arthur C. Graesser

b. Telephone Number: (901) 678-2742

c. ONR Program Officer: Jan Dickison

d. Program objective:

The goal of this project is to investigate how adults reason about everyday devices (e.g., locks, doorbells, dishwashers, clutches) when the devices break down. Deep comprehension of a device is presumably manifested by reasoning in the face of the malfunctioning device, because the comprehender has to diagnose the fault and figure out how to correct it. The primary hypothesis is that good comprehenders will ask and answer good questions. A good question is defined as a question that identifies a plausible fault that causally explains the breakdown. A bad question addresses irrelevant content that does not provide a likely explanation of the malfunction.

We collected data from 108 college students at the University of Memphis. The participants read 6 illustrated texts on everyday devices: a cylinder lock, an electronic bell, a car temperature gauge, a clutch, a toaster, and a dishwasher. The device mechanisms were extracted from a book with illustrated texts, The Way Things Work (Macaulay, 1988). After reading about each device, the participants subsequently received scenarios in which the device breaks down (e.g., in the context of a cylinder lock, suppose that the key turns, but the bolt does not move). During this time, they are asked either to “think aloud” or to generate questions for two minutes while attempting to diagnose and repair the malfunctions. We examined the questions that learners ask and answer when they are given the breakdown scenario. We measured the volume and quality of the content produced during the question asking or think aloud task. The content of the questions was also mapped onto conceptual graph structures (e.g., semantic networks, knowledge structures) for the illustrated texts on devices; these structures include component hierarchies, spatial region hierarchies, causal chains/networks, goal/plan/action hierarchies, and property descriptions that are depicted in either text or picture form.

After providing the question asking and think aloud protocols for all 6 devices, the participants completed an objective test on their understanding of the devices. This consisted of six 3-alternative, forced-choice questions about each device (36 total questions). Following this objective test of device comprehension, participants completed a battery of tests that measured their cognitive abilities and personality. The tests of cognitive ability include the ASVAB (the Armed Services Vocational Aptitude Battery), working memory span, spatial reasoning, and exposure to print. Male and female participants also completed the NEO inventory, which measures individuals on the “big five” personality factors: neuroticism, extroversion, openness, agreeableness, and
conscientiousness. We examined whether these measures of individual differences could predict comprehension scores and measures of question asking.

e. Significant Results during the Last Year

The results of the study supported the hypothesis that deep comprehenders of devices ask good questions when confronted a breakdown scenario. Mechanical comprehension and gender were the two most robust predictors of both the device comprehension scores (predicting 44% of the variance) and the quality of questions in the question asking task (predicting 38% of the variance) when multiple regression analyses were conducted. The profile of correlations with question quality were nearly identical with the profile of correlations with device comprehension scores. Question quality is therefore an excellent litmus test of how well a college student understands the mechanisms underlying a device. The volume of content in the question asking task and the think aloud task is a poor index of deep comprehension. Similarly, the quality of content in the think aloud task is also not a good index of deep comprehension. The questions asked by students with high mechanical comprehension scores had two characteristics: (a) the questions converged on components in the mechanism that are plausible faults and (b) the questions had a more fine-grained elaboration of the parts, processes, and relations that specified how the breakdown occurred. Stated differently, there was high convergence and high mechanistic detail.

f. Summary of Plans for Next Year’s Work

Experiments will be conducted that collect eye tracking data from college students while they read illustrated texts about everyday devices, and while they generate questions when these devices break down. Participants with high mechanical comprehension are predicted to have their eye fixations concentrate on faulty components and processes whereas those with low mechanical reasoning should have more undiscriminating patterns of eye fixations. More generally, we will relate the patterns of eye tracking to elements in the illustrated texts, to nodes in the conceptual graph structures, and to measures of individual differences. We will investigate the cognitive mechanisms that explain question asking and eye fixations in the context of a breakdown scenario. There is a more practical implication: A quick way to find out whether a sailor has a talent for understanding or operating a device is to present a break down scenario and to record the sailor’s question asking behavior and eye fixations.

g. Names of Graduate Students Currently Working on the Project:

Victoria Pomeroy
Brent Olde
Shulan Lu (starting July, 1999)
PART III

These are the three view graphs: An introductory 5-part viewgraph and two supporting viewgraphs. The paragraph descriptions are below. A Powerpoint file with the graphs is in a separate file.

VIEWGRAPH 1

The goal of this project is to investigate how adults reason about everyday devices (e.g., locks, doorbells, dishwashers, clutches) when the devices break down. Deep comprehension of a device is presumably manifested by reasoning in the face of the malfunctioning device, because the comprehender has to diagnose the fault and figure out how to correct it. The primary hypothesis is that good comprehenders will ask and answer good questions. A good question is defined as a question that identifies a plausible fault that causally explains the breakdown. A bad question addresses irrelevant content that does not provide a likely explanation of the malfunction. We collected data from 108 college students at the University of Memphis. The participants read 6 illustrated texts on everyday devices: a cylinder lock, an electronic bell, a car temperature gauge, a clutch, a toaster, and a dishwasher. The device mechanisms were extracted from a book with illustrated texts, *The Way Things Work* (Macaulay, 1988). After reading about each device, the participants subsequently received scenarios in which the device breaks down (e.g., in the context of a cylinder lock, suppose that the key turns, but the bolt does not move). During this time, they are asked either to “think aloud” or to generate questions for two minutes while attempting to diagnose and repair the malfunctions. We examined the questions that learners ask and answer when they are given the breakdown scenario. We measured the volume and quality of the content produced during the question asking or think aloud task. These were correlated with scores on a device comprehension test and 20 measures of individual differences.

VIEWGRAPH 2

The results of the study supported the hypothesis that deep comprehenders of devices ask good questions when confronted a breakdown scenario. Mechanical comprehension and gender were the two most robust predictors of both the device comprehension scores (predicting 44% of the variance) and the quality of questions in the question asking (QA) task (predicting 38% of the variance) when multiple regression analyses were conducted. The profile of correlations with question quality were nearly identical with the profile of correlations with device comprehension scores. Question quality is therefore an excellent litmus test of how well a college student understands the mechanisms underlying a device. The volume of content in the question asking task and the think aloud (TA) task was a poor index of deep comprehension. Similarly, the quality of content in the think aloud task is also not a good index of deep comprehension.
VIEWGRAPH 3

The content of the questions was also mapped onto conceptual graph structures (e.g., semantic networks, knowledge structures) for the illustrated texts on devices; these structures include component hierarchies, spatial region hierarchies, causal chains/networks, goal/plan/action hierarchies, and property descriptions that are depicted in either text (T) or picture form (P). The questions asked by students with high mechanical comprehension scores had two characteristics: (a) the questions converged on components in the mechanism that are plausible faults and (b) the questions had a more fine-grained elaboration of the parts, processes, and relations that specified how the breakdown occurred. Stated differently, there was high convergence and high mechanistic detail. In the Powerpoint figure, the nodes with red dots are the likely faults for the breakdown scenario (The key turns but the bolt does not move).
Question-driven Explanatory Reasoning About Devices that Malfunction

OBJECTIVE

To assess whether an adult has a deep comprehension of a device.
To explore the cognitive mechanisms that underlie deep comprehension of illustrated texts.
To develop models of human question asking and answering

APPROACH

Give breakdown scenarios after college students read illustrated texts from *The Way Things Work*.
Collect question asking and think aloud protocols in the context of a breakdown scenario.
Collect measures of cognitive ability and personality.

ACCOMPLISHMENTS

Participants with high mechanical comprehension ask good questions.
Good questions converge on faults and dissect processes.

TRANSITIONS

Next project collects eye tracking data.
Patterns of eye movements should quickly manifest deep comprehension.
CORRELATIONS

<table>
<thead>
<tr>
<th>PREDICTOR</th>
<th>DEVICE</th>
<th>QUALITY</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>COMPR.</td>
<td>QA</td>
</tr>
<tr>
<td>Spatial reasoning</td>
<td>.46</td>
<td>.42</td>
</tr>
<tr>
<td>Working memory</td>
<td>.01</td>
<td>.04</td>
</tr>
<tr>
<td>Print Exposure (ART)</td>
<td>.27</td>
<td>.30</td>
</tr>
<tr>
<td>ASVAB (g)</td>
<td>.59</td>
<td>.41</td>
</tr>
<tr>
<td>Mechanical comprehension</td>
<td>.63</td>
<td>.56</td>
</tr>
<tr>
<td>Electronics</td>
<td>.56</td>
<td>.52</td>
</tr>
<tr>
<td>General science</td>
<td>.60</td>
<td>.48</td>
</tr>
<tr>
<td>Auto &amp; Shop</td>
<td>.52</td>
<td>.40</td>
</tr>
<tr>
<td>Mathematics knowledge</td>
<td>.56</td>
<td>.45</td>
</tr>
<tr>
<td>Arithmetic reasoning</td>
<td>.52</td>
<td>.37</td>
</tr>
<tr>
<td>Numerical operations</td>
<td>.12</td>
<td>.10</td>
</tr>
<tr>
<td>Word knowledge</td>
<td>.42</td>
<td>.29</td>
</tr>
<tr>
<td>Paragraph comprehension</td>
<td>.27</td>
<td>.11</td>
</tr>
<tr>
<td>Coding speed</td>
<td>.08</td>
<td>-.08</td>
</tr>
<tr>
<td>Gender</td>
<td>.37</td>
<td>.37</td>
</tr>
<tr>
<td>Age</td>
<td>.01</td>
<td>.01</td>
</tr>
<tr>
<td>Openness (NEO)</td>
<td>.32</td>
<td>.24</td>
</tr>
<tr>
<td>Neuroticism</td>
<td>-.08</td>
<td>.04</td>
</tr>
<tr>
<td>Extroversion</td>
<td>-.01</td>
<td>-.03</td>
</tr>
<tr>
<td>Agreeableness</td>
<td>-.03</td>
<td>-.12</td>
</tr>
<tr>
<td>Conscientiousness</td>
<td>-.01</td>
<td>-.04</td>
</tr>
</tbody>
</table>

\( r = .19 \) is significant at \( p < .05 \).
# QUESTIONS ABOUT A CYLINDER LOCK

**Mechanical Comprehension**

<table>
<thead>
<tr>
<th>Question</th>
<th>HIGH</th>
<th>LOW</th>
</tr>
</thead>
<tbody>
<tr>
<td>Is it the right key?</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>What kind of lock is it?</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Is the spring broken?</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Does spring keep bolt from moving?</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>Is the spring pulling back the bolt?</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Is the spring making the bolt get stuck?</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Is the cam broken?</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>Is the cam moving?</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Is the cam moving back the bolt?</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Is the bar that fits under cam broken?</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>Is the cam disconnected/out-of-synch with the cylinder?</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>Is the cylinder turning?</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Is the cylinder turning the cam?</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Is the bolt stuck in the slot?</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Is the bolt connected to the bar?</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Are the pins broken?</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>Do the pins lift right?</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>What are the pins used for?</td>
<td>0</td>
<td>2</td>
</tr>
</tbody>
</table>

---

Number if participants 11 11