Question Answering in the Context of Illustrated Expository Text

William B. Baggett
Department of Psychology and the
Institute for Intelligent Systems
The University of Memphis
Memphis, TN 38152
baggettw@cc.memphis.edu

Arthur C. Graesser
Department of Psychology and the
Institute for Intelligent Systems
The University of Memphis
Memphis, TN 38152
a-graesser@memphis.edu

Abstract

We investigated how college students answer questions about the content of illustrated expository text. Subjects studied illustrated texts describing causal event chains that unfold during the operation of everyday machines. Subjects subsequently provided written answers to questions about events occurring in each machine. Four types of questions were asked: why did event X occur?, how did X occur?, what are the consequences of X occurring?, and what if X didn't occur?. In our analysis of the answer protocols, we adopted the theoretical framework of the QUEST model of human question answering (Graesser & Franklin, 1990). The present study supported predictions generated from three components of the QUEST model: question categorization, utilization of information sources, and convergence principles. Our results also revealed two novel findings. First, subjects had a bias toward sampling information from the text more than from the picture. Second, subjects tended to sample information depicted in both the text and the picture.

Introduction

How do adults answer deep-reasoning questions about the inner workings of machines after studying texts with illustrations? Figure 1 presents an illustrated text describing the workings of a cylinder lock. We asked college students to study these materials at their leisure and to provide written answers to deep explanatory questions such as the following:

Why does the cylinder rotate?
How does the cylinder rotate?
What are the consequences of the cylinder rotating?
What if the cylinder did not rotate?

In our analysis of the answer protocols, we adopted the theoretical framework of the QUEST model of human question answering (Graesser & Franklin, 1990). The present study is the first to evaluate QUEST in the context of illustrated expository texts.

The role of illustrations in text comprehension has recently received considerable attention (Glenberg & Langston, 1992; Hegarty & Just, 1993; Mandl & Levin, 1989; Mayer & Anderson, 1989; Schnotz & Kulhavy, 1994). However, these studies have not directly investigated how deep questions are answered in the context of the illustrated texts. Given that answers to deep explanatory questions are an important reflection of understanding (Graesser, Singer, & Trabasso, 1994; Schank, 1986), the present study investigated the mechanisms that generate answers to these questions.

QUEST: A Model of Human Question Answering

We focused on three components of the QUEST model: question categorization, information sources, and convergence principles.

Question Categorization

QUEST assumes (a) that there is a catalog of question categories, (b) that each question category has an associated question-answering procedure which operates on relevant knowledge structures to generate appropriate answers, and (c) that an incoming question is assigned to one of these question categories. The present study investigated four question stems: why, how, what-are-the-consequences, and what-if-not.

Information Sources

QUEST specifies the information sources (i.e., knowledge structures) that are active in working memory after a query occurs. We assumed that the text and the picture were the primary information sources available to subjects during question answering.

QUEST uses *conceptual graphs structures* (CGS's) to represent information from both the text and the picture (Baggett, 1994). This representational system is similar to those of conceptual dependency theory (Schank & Abelson, 1977) and causal chain theories (Trabasso, van den Broek, & Suh, 1989). A CGS consists of a group of statement *nodes* that are related by labeled, directed *arcs*. Figure 2 gives a portion of the CGS associated with the cylinder lock materials. In general, a statement node contains a predicate (e.g., verb, adjective, or adverb) and one or more arguments (e.g., noun or an embedded proposition). That is, a statement node is essentially a proposition that is assigned to a category (Kintsch & van Dijk, 1978). The CGS in Figure 2 contains 14 nodes that represent 8 causally related events, 4 supporting states of the world, and 2 goals of the user of the device.

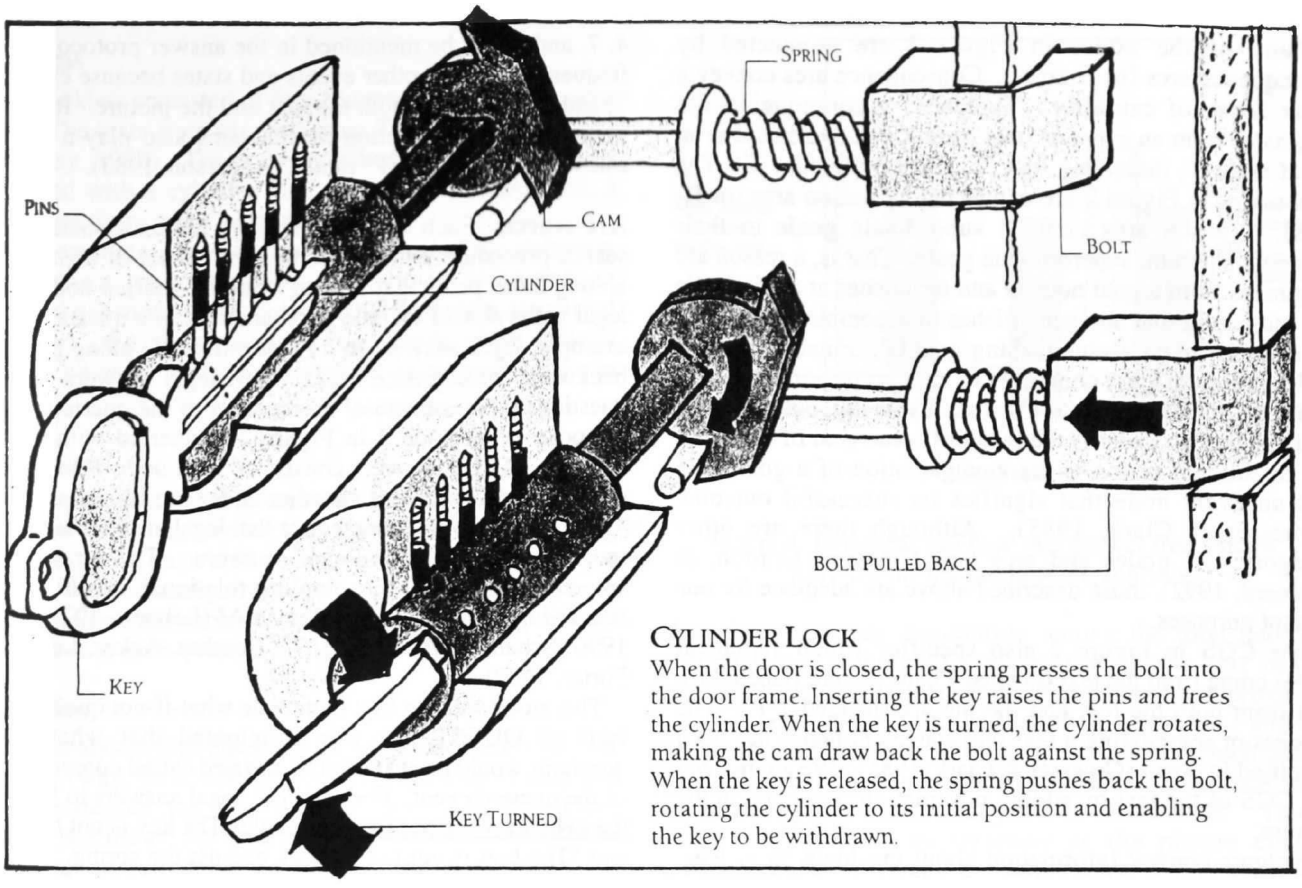

Figure 1: Example illustrated text describing a cylinder lock. (From *The Way Things Work* by David Macaulay. Compilation copyright (c) Dorling Kindersley Ltd., London. Illustration copyright (c) 1988 David Macaulay. Text copyright (c) 1988 David Macaulay, Neil Ardley. Reprinted by permission of Houghton Mifflin Company. All rights reserved.)

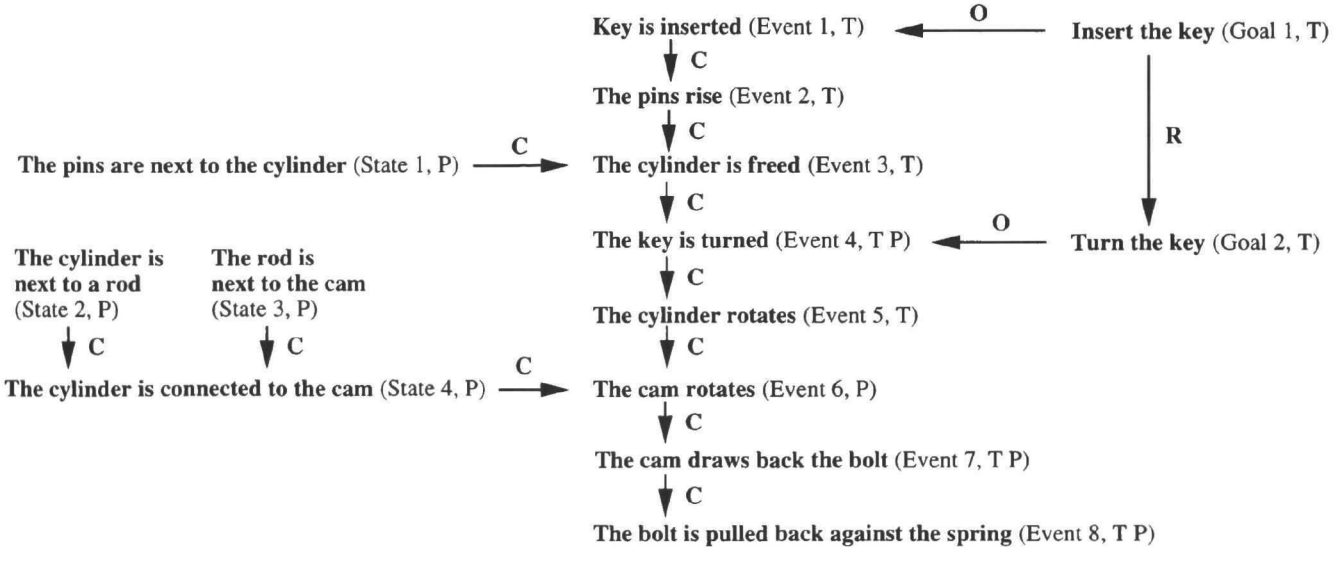

Figure 2: A portion of the cylinder lock materials represented as a conceptual graph structure. The arc categories include consequence (C), reason (R), and outcome (O). Each node is labeled as being depicted in the picture alone (P), the text alone (T), or both text and picture (T P).

Many of the nodes in Figure 2 are connected by consequence arcs (or C-arcs). Consequence arcs convey a weak sense of causality. Typically, a consequence arc originates from an event or state node X and terminates at an event node Y, indicating that X causes or enables Y. The goal nodes in Figure 2 are connected by reason arcs (or R-arcs). Reason arcs connect subordinate goals to their immediate parent, superordinate goals. That is, a reason arc originates from a goal node G and terminates at a goal node G', indicating that an agent wishes to accomplish goal G in order to facilitate accomplishing goal G'. Finally, the goal nodes in Figure 2 are connected to their corresponding event nodes by outcome arcs (or O-arcs). Typically, outcome arcs indicate that an event is the result of some goal of an agent. An intentional action is the amalgamation of a goal node and an event node that signifies its successful outcome (Graesser & Clark, 1985). Although there are other categories of nodes and arcs (see Graesser, Gordon, & Brainerd, 1992), those described above are adequate for our current purposes.

The CGS in Figure 2 also specifies which statement nodes come from the text (labeled *T*), the picture (labeled *P*), and from both the text and the picture (labeled *T P*). The process of segregating a text into statement nodes has been specified in detail (Graesser & Clark, 1985). We augmented the CGS of the text by adding information contained in the picture.

Pictures convey information about machines in at least two ways. First, spatial adjacencies of machine components convey "next-to" relations. For example, the picture in Figure 1 includes the information that "the pins are next to the cylinder." Thus, a state node corresponding to this relation (state node 1) was added to the CGS. Second, pictures convey information about the movements of some components through the use of arrows. For example, an arrow on the second depiction of the system in Figure 1 indicates the event, "the cam rotates." Thus, a corresponding event node (event node 6) was added to the CGS.

Convergence Principles

Convergence principles narrow down the set of nodes which serve as good answers to a question. The quality of answers was assumed to be positively correlated with the frequency of the answers produced in the question answering protocols. The convergence principles of QUEST include (1) the intersection principle, (2) the arc search procedures, and (3) structural distance.

Intersection principle. The intersection principle predicts that the quality of a potential answer node increases with the number of information sources that store the node. That is, nodes which come from both the text and the picture are predicted to be better answers than those coming from the text alone or from the picture alone. Suppose event node 3 in Figure 2 were queried with a consequence question, i.e., What are the consequences of the cylinder being freed?. The intersection principle predicts that events 4, 7, and 8 will be mentioned in the answer protocols more frequently than the other events and states because events 4, 7, and 8 come from both the text and the picture. It should be noted that intersection mechanisms also play a critical role in Anderson's ACT* model (Anderson, 1983).

Arc search. Each question category has an associated arc search procedure which specifies the subset of *legal* paths among those paths that radiate from the queried node. The legal paths should include good answers to a question. For example, legal answers to how questions lie along paths of backward consequence arcs. Thus, legal answers to how questions represent causal antecedents of the queried event. Suppose event node 3 in Figure 2 is queried with a how question. Legal answers consist of state node 1 and event nodes 1 and 2, whereas all other nodes are illegal answers. The arc search principle predicts that legal answers are given more frequently that illegal answers. The strategy of answering questions by sampling relational arcs has been adopted by a number of models in AI (Lehnert, 1978; Ram, 1990; Schank & Ableson, 1977; Souther, Acker, Lester, & Porter, 1989).

This study was the first to include what-if-not questions in tests of QUEST. It was anticipated that what-if-not questions would be answered by negated causal consequences of the queried event. For example, legal answers to What if the cylinder were not freed? include "The key is not turned," and "The bolt is not pulled back against the spring." Legal answers to consequence questions consist of causal consequences of the queried event. Graesser & Hemphill (1991) found that legal answers to why questions consist of both causal antecedents and causal consequences of the queried event. Subjects apparently developed teleological representations of mechanisms from the biological and technological sciences, as evidenced by their tendency to answer why questions in those domains with causal consequences.

Structural distance. The structural distance between two nodes is the number of arcs on the shortest path between the nodes. QUEST predicts that answer quality decreases as the structural distance between the queried node and the potential answer node increases. According to this principle, when event node 3 in Figure 2 is queried with a how question, event node 2 is a better answer than event node 1. This structural distance gradient is compatible with the spreading activation mechanism in the ACT* model (Anderson, 1983). Note that the predicted effect of structural distance may be challenged by the presence of illustrations in the instructional materials. That is, illustrations may enhance subject's memories for portions of the causal sequence, which allows subjects to make more distant associations.

Methods

Subjects

The subjects were 32 undergraduates from The University of Memphis.

Materials

Study booklets consisted of four illustrated texts on separate pages. The texts were excerpts from *The Way Things Work* (Macaulay, 1988). The texts described causal event chains associated with a cylinder lock, a car engine crankshaft, a microwave oven, and a water meter.

From each text, four events were selected for querying. Each subject read all of the texts, but a given subject answered questions about only one of the events associated with a text. We counterbalanced the order in which subjects read the passages and the assignment of queried events to subjects.

Procedure

Subjects were instructed to study the illustrated texts at their own pace before answering any questions. After reading each text, they answered four questions about the text. Subjects were encouraged to refer back to the illustrated texts during question answering and to fill the three lines provided for answers to each question.

Variables

For all analyses, the independent variables were question category (why, how, consequence, versus what-if-not) and answer category (as described below). The dependent variable was the answer frequency per question. Most (79%) of the answers referred to explicit information so we did not analyze inferences. In the analyses of information sources and structural distance, each empirical answer frequency was compared to an appropriate base rate answer frequency. The base rate frequencies reflect random sampling of the nodes in the illustrated texts. The base rate frequency for each answer category was computed by multiplying the mean number of explicit answers statements subjects produced per question, 1.7, by the proportion of CGS statement nodes in that category.

Results & Discussion

Written question-answering protocols were analyzed along three dimensions specified by the QUEST model of human question answering: information sources, arc search procedures, and structural distance. All comparisons discussed in this section were statistically reliable, unless stated otherwise.

Information Sources

The explicit answers were segregated into those coming from the text alone, the picture alone, versus both. Figure 3 summarizes the results of the information source analysis. For each answer category (text alone, picture alone, versus both), the empirical answer frequency was compared to the base rate answer frequency for that category. The base rate frequencies reflect random sampling of the nodes in the illustrated texts.

One of the provocative findings is that subjects gave answers which come from the text alone much more often

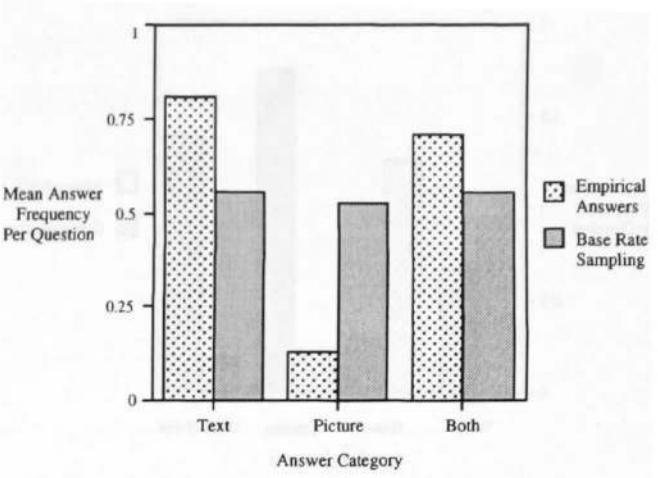

Figure 3: Empirical and base rate answer frequencies from the information source analysis.

than their base rate availability among the total pool of information. Apparently, information in the text has a privileged status. Subjects focused on information provided in the text alone during encoding, retrieval, and/or articulation. In contrast, subjects gave answers which come from the picture alone much less often than their base rate availability among the total pool of information. Thus, when information was presented in the picture alone, subjects tended to ignore the information during study or filter it out during question answering.

Another provocative finding is that subjects gave answers which come from both the text and the picture more often than their base rate availability. It should be noted that this base rate includes an adjustment for the double presentation of the information. Why do subjects focus on information presented in both the text and the picture over and above its double presentation? One explanation of this effect is QUEST's intersection principle. According to QUEST, a statement node represents a better answer to a question to the extent that it comes from multiple information sources. For example, a statement that is read in both a newspaper article and a book comes from at least two explicit information sources. From this point of view, the simple fact that the text and the picture are distinct information sources causes them together to have a greater than additive impact on answer frequency.

An alternative explanation of the latter finding comes from the dual code hypothesis (Paivio, 1971). According to the dual code hypothesis, memories are more likely to be recalled to the extent that they are tied to both the visual and verbal modalities.

In summary, subjects are biased toward sampling information from the text more than from the picture. This finding is compatible with the results of eye tracking studies conducted by Hegarty and Just (1993). These studies indicate that it is the text rather than the picture that drives comprehension. Moreover, subjects exhibit a bias toward information coming from both the text and the picture. That is, the effect of presenting information in the text can be strengthened by presenting it in the picture as well. The

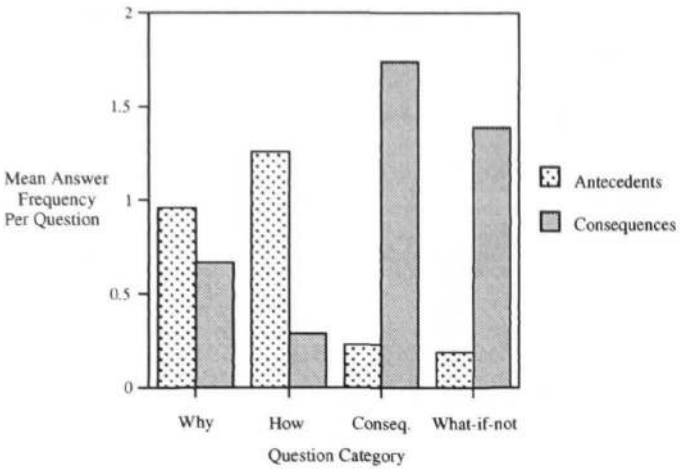

Figure 4: Empirical answer frequencies from the arc search analysis.

latter effect is compatible with QUEST's intersection principle and with the dual code hypothesis.

Arc Search Procedures

The explicit answers were segregated into causal antecedents versus causal consequences of the queried event. Figure 4 summarizes the results of the arc search analysis. For each question type, we compared the empirical answer frequencies of causal antecedents and causal consequences of the queried event.

The data clearly substantiate the predictions made by QUEST's arc search procedures. The arc search component correctly predicted that the legal answers for how, consequence, and what-if-not questions sample causal antecedents, causal consequences, and negated causal consequences of the queried event, respectively. QUEST accommodates both causal antecedents and causal consequences as answers to why questions, and in fact the difference in answer frequencies was not significant.

Structural Distance

The explicit answers were categorized according to the number of arcs on the shortest path between the queried node and the answer node (one, two, versus three or more). Figure 5 summarizes the results of the structural distance analysis. For each question type and each answer category, the empirical answer frequency was compared to the appropriate base rate answer frequency. We focus on the frequencies of answers which are either near (one arc) or distant (three or more arcs) from the queried node. Again, the base rate frequencies reflect random sampling of the nodes in the illustrated texts.

The data show some support for the predicted structural distance gradient. When subjects answered what-if-not questions, they apparently sampled the nodes on legal paths randomly with regard to structural distance. For all comparisons involving why, how, and consequence questions, however, the differences were in the direction predicted by QUEST's structural distance gradient. That is, the frequencies of answers which were near (one arc away from) the queried event node were higher than their

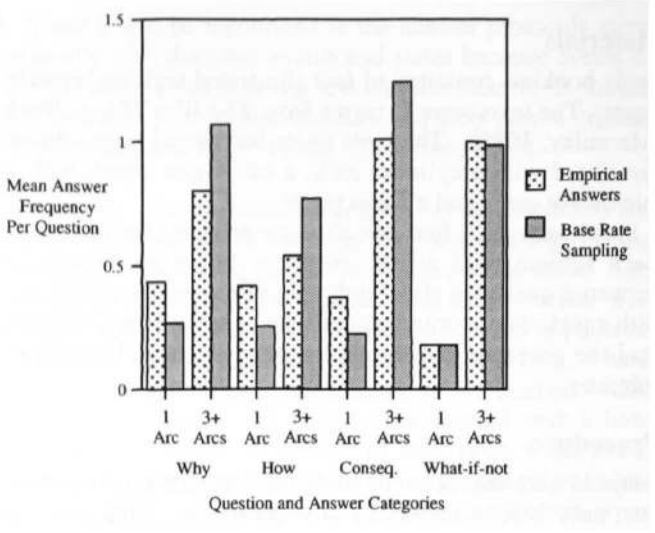

Figure 5: Empirical and base rate answer frequencies from the structural distance analysis.

respective base rates; the frequencies of answers which were distant (three or more arcs away) from the queried event node were lower than their respective base rates. These results are consistent with the prediction that answer quality decreases as the structural distance between the queried node and the potential answer node increases.

References

Anderson, J. R. (1983). *The architecture of cognition.* Cambridge, MA: Harvard University Press.

Baggett, W.B. (1994). Question Answering and Illustrated Expository Text. Unpublished master's thesis. Memphis, TN: The University of Memphis.

Glenberg, A.M., & Langston, W.E. (1992). Comprehension of illustrated text: Pictures help to build mental models. *Journal of Memory and Language*, 31, 129-151.

Graesser, A.C., & Clark, L.C. (1985). *Structures and procedures of implicit knowledge.* Norwood, NJ: Ablex.

Graesser, A.C., & Franklin, S.P. (1990). QUEST: A cognitive model of question answering. *Discourse Processes*, 13, 279-303.

Graesser, A.C., Gordon, S.E., & Brainerd, L.E. (1992). QUEST: A model of question answering. *Computers and Mathematics with Applications*, 23, 733-745.

Graesser, A.C., & Hemphill, D. (1991). Question answering in the context of scientific mechanisms. *Journal of Memory and Language*, 30, 186-209.

Graesser, A.C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review*, 101, 371-95.

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

Kintsch, W., & van Dijk, T.A. (1978). Toward a model of text comprehension and production. *Psychological Review*, 85, 363-394.

Lehnert, W.G. (1978). *The process of question answering.* Hillsdale, NJ: Erlbaum.

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

Mandl, H., & Levin, J.R. (1989). *Knowledge acquisition from text and pictures.* Amsterdam: North Holland.

Mayer, R.E., & Anderson, R.B. (1989). Models for understanding. *Review of Educational Research, 59,* 43-64.

Paivio, A. (1971). *Imagery and verbal processes.* New York: Holt, Rinehart, and Winston.

Ram, A., (1990). Knowledge goals: A theory of interestingness. In *Proceedings of the 12th Annual Conference of the Cognitive Science Society* (pp. 206-214). Hillsdale, NJ: Erlbaum.

Schank, R.C., Abelson, R. (1977). *Scripts, plans, goals and understanding: An inquiry into human knowledge structures.* Hillsdale, NJ: Erlbaum.

Schnotz, W., & Kulhavy, R. (1994). *Comprehension of graphics.* Elmsford, NY: Pergamon Press.

Souther, A., Acker, L., Lester, J., & Porter, B. (1989). Using view types to generate explanations in intelligent tutoring systems. *Proceedings of the 11th Annual Conference of the Cognitive Science Society* (pp. 123-130), Hillsdale, NJ: Erlbaum.

Trabasso, T., van den Broek, P.W., & Suh, S.Y. (1989). Logical necessity and transitivity of causal relations in stories. *Discourse Processes,* 12, 1-25.