Designing Hypermedia Manuals to Explain how Machines Work: Lessons from Evaluation of a Theory-based Design [1]

Mary Hegarty [2]
Jill Quilici[2]
N. Hari Narayanan[3]
Selma Holmquist[3]
Roxana Moreno[2]

Technical Report CSE98-04

April 3, 1998

Visual Information, Intelligence & Interaction Research Group
Department of Computer Science & Engineering
Auburn University
Alabama 36849-5347
USA
http://www.eng.auburn.edu/cse/research/vi3rg/vi3rg.html

---

[1]Paper to be presented at *Hyped-Media to Hyper-Media: Toward Theoretical Foundations of Design, Use and Evaluation,* A Workshop at CHI'98, The ACM SIGCHI Conference on Human factors in Computing Systems, Los Angeles, April 1998.

[2] Department of Psychology, University of California, Santa Barbara, CA 93106
[3] Department of Computer Science & Engineering, Auburn University, Auburn, AL 36849

1. Introduction

Recent technological advances in multimedia capabilities present a tantalizing array of choices to developers of instructional materials. For example, we can now present text visually on a computer screen or as an audio clip, we can present static or animated diagrams, and we can allow users to navigate through hypermedia documents in any order, rather than being constrained to the linear order of the print medium. On the one hand there are many unquestioned beliefs about the power of multimedia in instruction, for example that diagrammatic representations are better than sentential representations, three dimensional representations are better than two-dimensional ones, animated diagrams are more effective than static diagrams and interactive graphics are better than non-interactive graphics (Scaife & Rogers, 1996). On the other hand, there are few empirically established guidelines on how to choose among the various capabilities for optimal design of hypermedia manuals. Therefore it is important to develop principled guidelines for the design of hypermedia systems and to test them empirically, documenting both the successes and failures of designs that embody the guidelines.

In this research project, we are studying the use of multimedia for the specific application of developing instructional hypermedia manuals that explain how machines work. Based on prior empirical and computational research (e.g., Hegarty, 1992; Mayer, 1989; Narayanan, Suwa & Motoda, 1994a, 1994b; Novak, 1995) we developed a theoretical model of stages in the comprehension of a machine from diagrams and accompanying text (Narayanan & Hegarty, in press). We then considered the implications of this model for hypermedia manual design by identifying sources of error at each stage in our comprehension model and enumerating design guidelines for ameliorating these. We designed and implemented a prototype hypermedia manual based on this model, and evaluated the model and the prototype in three experiments. Results from these experiments appear to indicate that unique characteristics of hypermedia - hyperlinking semantically related information and dynamic presentations such as animations coupled with audio - do not provide a significant comprehension and reasoning advantage over static multimodal (text and diagrams) presentations when both types of presentations are designed according to the theoretical model.

2. A Model of Diagram Comprehension

We begin with a condensed description of the theoretical model upon which the initial hypermedia manual design was based. See (Narayanan & Hegarty, in press) for details. According to our model, comprehension of text-and-diagram descriptions of machines involves the following processes. We list them in order, although we do not propose that they are necessarily always accomplished in this order:

A. Machine Decomposition by Diagram Parsing. Diagrams of mechanical systems are made up of elementary shapes, such as rectangles, circles and cylinders, which represent objects such as pistons, gears and tubes. The first step in comprehension is to parse the connected diagram into these elementary shapes, i.e., units that correspond to objects. This process is analogous to identifying discrete words and clauses in a continuous speech sound and probably relies largely on perceptual mechanisms of object recognition (Biederman, 1987, Marr & Nishihara, 1978).

B. Constructing a Static Mental Model by Making Representational Connections The second stage in multi-modal comprehension involves making appropriate connections in memory among the units identified in Stage 1. This stage involves making two types of connections: (1) connections to prior knowledge and (2) connections to the representations of other machine components (Mayer & Sims, 1994).

(1) Connections to Prior Knowledge. First, the user must identify the components, that is, make connections between the diagrammatic elements identified at Stage 1 and their real-world

referents -- a process analogous to lexical access in language comprehension. For example, the user might represent that a rectangle represents a piston or a circle represents a gear. Prior knowledge can also provide additional information about components, such as what these are typically made of and if they are rigid or flexible. This information is valuable in making inferences about how components move and constrain each others behaviors.

(2) Connections to the representation of other machine components. Second, the user must represent the spatial relations between different machine components by building connections between the representations of these components. In understanding how a machine works, information about the spatial relations between mechanical components forms the basis for inferences about the motions of components, because these spatial relations determine how components affect and constrain each others' motions. Knowledge of spatial relations also aids in guiding the reasoning process along the chain of causality in the machine (Hegarty, 1992; Narayanan, Suwa & Motoda, 1994b).

C. Making Referential Connections. When diagrams are accompanied by text, an additional stage in comprehension is that of resolving coreference between the two media, i.e., making referential links between a noun phrase in the text (e.g., "the piston") and the diagrammatic unit that depicts its referent (e.g., a rectangle) (Novak, 1995). This step is crucial to constructing an integrated representation of the common referent of the text and diagram in memory as opposed to separate surface-level representations of the text and diagram. Making referential connections is also a necessary process when users have to integrate information in two different pictorial displays of the same machine, e.g., a schematic diagram and a realistic picture or photograph of the machine, or to construct a 3-dimensional representation from diagrams showing different perspective views of the same system.

D. Determining the Causal Chain of Events. When asked to predict the behavior of machines from static diagrams, people tend to reason about machine operation along the direction of causal propagation in the machine (Hegarty, 1992; Narayanan, Suwa & Motoda, 1994b). Therefore, we hypothesize a fourth stage of comprehension that involves identifying the potential causal chains of events in the operation of the machine, or "lines of action" in the machine.

E. Constructing a Dynamic Mental Model by Mental Simulation and Rule-based Inference. The final stage of comprehension is that of constructing a dynamic mental model of the machine by inferring and integrating the behaviors of individual components. We refer to this process as *mental animation*. Cognitive and computational models (Hegarty, 1992; Narayanan, Suwa & Motoda, 1994a; 1994b) suggest that this is an incremental process in which the reasoner considers the components or subsystems individually, assesses the influences acting on each, infers the resulting behavior of each, and then proceeds to consider how this behavior affects the next component or subsystem in the causal chain.

3. Initial Design of a Hypermedia Manual

In a recent paper (Narayanan & Hegarty, in press) we discussed how hypermedia manuals might be designed to facilitate users in accomplishing the five stages of comprehension outlined above. For each stage in the comprehension model, we identified potential sources of comprehension error that users might encounter, and developed hypermedia design guidelines intended to ameliorate these difficulties. In this section we outline these guidelines and explain how they were applied to develop a prototype hypermedia manual that explains how a toilet tank (a flushing cistern) works. Though this is a device that is very familiar in use to people, its inner workings are not intuitively obvious. It is a relatively complex device that has two main subsystems, a water output system that flushes water into the toilet tank and a water inlet system that refills the tank for the next use. Explaining a toilet tank presents interesting challenges for our theory because its operation involves two causal chains of events that occur in tandem but are also temporarily dependent on each other. The particular toilet tank explained in our manual (which is shown in Figure 1) also contains a siphon, raising the interesting question of how to explain this basic physics principle in the context of explaining how a specific machine works.

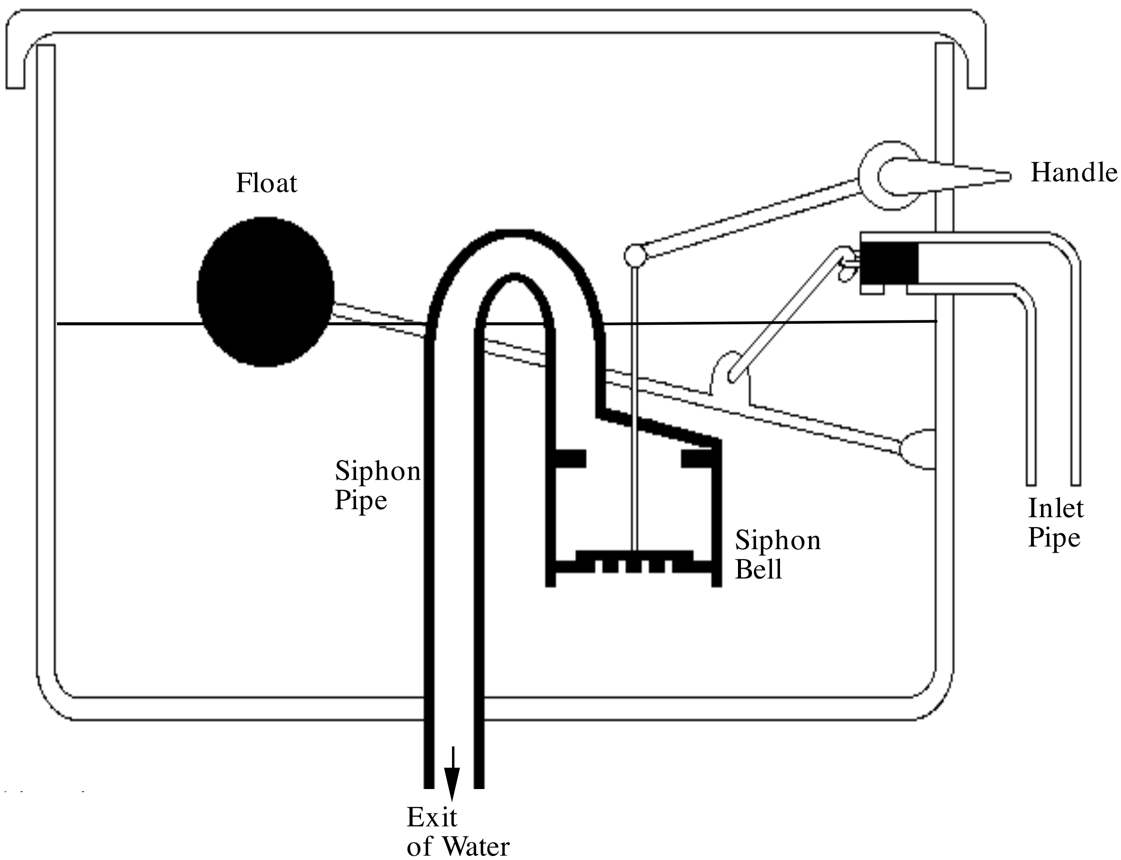

Figure 1: A Flushing Cistern

The hypermedia manual contains seven sections designed to guide users through the stages of comprehension in our model. Presentations in each section can contain multiple media: static text, audio narratives, static diagrams and animations. A labeled schematic diagram provides continuity across sections, since it appears in all sections. All diagrams are consistently labeled with terms appearing in the text. Animations are accompanied by synchronized audio narratives. Textual descriptions may contain words and phrases in hypertext that the user can click on to get more information.

Section 1. The primary objectives of this section are to help the user in decomposing the machine into its components and building referential connections between the text and diagram. Previous research has pointed out that diagrams are often underspecified in that they do not contain enough information for a user to identify whether two or more connected units in a diagram represent separate objects or parts of a single object (Novak, 1995). For example, a diagram element, such as a line, might represent the edge of an object, or an object itself (a rope). Furthermore, resolving coreferences between text and graphics can be a source of comprehension difficulty (Mayer, 1989; Mayer & Sims, 1994). To facilitate users in these processes, our manual the first section presents a cross-sectional diagram of the toilet tank, in which decomposition is facilitated by labeling the different functional components, by presenting them in different colors, and by allowing users to click on the label of any component and have this highlighted in the diagram.

Section 2: The decomposition of a mechanical system can sometimes be hierarchical, such that the system breaks down into *functional* subsystems, which can themselves be broken down into more elementary components. The objective of Section 2 is to facilitate identifying the functional subsystems to which the components belong -- a water output system that flushes water into the toilet tank and a water input system that refills the tank. The first presentation in this section is a schematic diagram with accompanying text that outlines the various subsystems of the machine tank. This presentation also allows the user to select an exploded view of the tank in which the input and output systems are separated in space. In two further presentations, the user views diagrams highlighting the components of first the output system and then the input system, each accompanied by text describing the function of the relevant subsystem.

Section 3:. The objective of section 3 is to facilitate construction of a static mental model of the machine. This includes identifying the real-world components to which diagram units refer, and retrieving prior knowledge about these components (what they are made of, their principles of operation, etc.) Highly schematized depictions that do not look like their real-world referents can hamper the process of identifying the depicted components. Furthermore, access to knowledge typically not available from a schematic diagram (e.g., whether components are rigid or flexible) is required for correct causal inferences (Narayanan, Suwa & Motoda 1994a, Schwartz & Black, 1996). In this section, the user is shown a cross-sectional view of the toilet tank and taken on a guided tour of the components. Text describes each of the components in turn, pointing out their linkages to other system components, and other information that is not visible in the diagram, such as their material composition and function. Only the text about one component is visible at a time, and to aid the construction of referential connections, the component in question is also highlighted in the diagram.

Section 4: This section is designed to encourage the user to reason about the causality and dynamics of the machine. Previous research has shown that the generation of ideas or self-explanations improves learning (Chi, deLeeuw, Chiu & LaVancher, 1994). Therefore one of our guidelines is that users should be encouraged to mentally animate the machine (i.e. attempt to predict its behavior) before they are shown animations of the causal chain and movement of the machine. Section 4 does this by presenting users with the static diagram from the previous sections of the manual, and a set of multiple choice questions in which users are asked to imagine that a component of the system is moving in a given way and have to predict how another component of the system will be moving. Users are given feedback on whether their answers are right or wrong and if the answer is wrong, the user is directed to note the correct answer in the later sections that describe the causal chain and the system operation.

Section 5: The purpose of this section is to help the user in understanding the propagation of causality in the entire machine. It contains an audio commentary describing the operation of the toilet tank, which describes causal propagation within and across the water output and water inlet subsystems. Synchronized with the commentary, the corresponding components and paths of causal propagation are highlighted in the static schematic diagram. An “explain principle” button allows users to access a description of the fundamental physics principle underlying the operation of the water output system - the siphon (see section 7).

Section 6: The objective of this section is to convey how the machine actually operates by describing and showing the movements of its components. An audio commentary sequentially describes how the components move. An animation of the schematic diagram continuously cycles through the behavior of the system during the commentary. This structure is expected to correspond well to the serialization effect we have observed in mental animation (Hegarty, 1992; Narayanan, Suwa & Motoda, 1994a, 1994b), while preserving the realism of simultaneous operation of multiple components in the visual part of the animation. At the end of this narrated animation, users have the option of replaying it, or viewing a faster running silent animation. An “explain principle” button again leads users to a description of how a siphon works.

Section 7: The purpose of this section is to explain a fundamental physics principle underlying the behavior of the toilet tank - the siphon. It contains a schematic diagram of the machine with the siphon bell and pipe highlighted. The text describes how a siphon works and how it applies to the operation of the water output system of the toilet tank. Optionally, users can also view a silent animation of the siphon effect. This section is not part of the sequential path through sections of the manual. Instead, it is reached optionally from sections 5 or 6, in response to a user clicking the “explain principle” button.

Navigation and Guidance

Inter-section Navigation: The model of machine comprehension outlined above suggests a sequence of stages of comprehension, such that the later stages are at least somewhat dependent on successful completion of the earlier ones. In particular, representation of the spatial relations between device components (Stage B(2)) is dependent on first decomposing the system into individual components (Stage A), and the stages of finding lines of action and constructing a kinematic mental model (Stages D and E) are dependent on successful construction of a static model of the machine (Stages B and C). Therefore, in this initial version of the manual, we constrained navigation, so that users studied the sections of the manual in the order of stages described in the model. Starting with Section 1, a user proceeds through sections 1 to 7 using an overall “map” of the hypermedia manual, which shows icons for all the sections of the manual and is color coded to show users their current place in the system, the sections that they have already studied and the sections that they are currently allowed to move to. The exception to this forward traversal is section 8, which can be accessed from Section 6 or 7, as described above. Backward traversal to presentations in previously visited sections is unrestricted. Once the user has seen a section, he or she may return to it at any time. The user does this by clicking on the relevant section in the map of the manual. Sections in the map are highlighted and become mouse sensitive based on the user’s history, so that clicking a section on this graphic will only be successful if the user has already visited this section.

Intra-section Navigation: Intra-section navigation facilities vary from fully unconstrained to fully constrained. Sections 1 and 2 allow fully unconstrained navigation. Users may view any part of the sections at any time. Navigation within Section 3 is semi-constrained. Users are first taken on a guided tour of the components one by one. After this tour is over, users can freely revisit the description of any of the components. Navigation within Section 4 is fully constrained in that users can only move sequentially through the questions. Navigation within Section 5 is also fully constrained in that the users are taken through the entire commentary of causal propagation within and across the subsystems. They however do have the option of replaying this narration. Section 6 provides semi-constrained navigation. Once the initial narrated animation has been viewed, users may replay it or view another silent animation. Navigation within Section 7 is fully unconstrained.

Our comprehension model and guidelines for hypermedia design were based largely on prior experiments conducted with novices. It is possible that the highly directive nature of the manual will not be optimal for more knowledgeable users. For example previous research (Hegarty & Just, 1993) showed that more knowledgeable readers are less text-directed in their processing of diagrams of simple mechanical systems. To gather preliminary data on how learning from hypermedia is affected by prior knowledge, we gathered information on participants’ previous physics training and practical experience with mechanical systems and assessed how these factors modulated learning from hypermedia.

4. Evaluation

Experiment 1

In Experiment 1, we compared learning from the hypermedia manual to learning from a paper-and-pencil printout of the text and diagrams used in the manual (which we will refer to as the full text condition), and learning from a paper-and-pencil description that just showed a labeled diagram of the toilet tank, which was accompanied by the text describing the movement of the components in order of the causal chain (the causal text condition). The main differences between the hypermedia manual and full text conditions were the absence of hyperlinks and animations in the full text condition and the presentation of the verbal description of the causal chain and movement of components, which was presented visually rather than auditorally. The main differences between the full text and causal text conditions were that the sections explicitly pointing out the subsystems of the toilet tank, the connections between components and the material composition of the components were not included in the causal text.

Method

Participants: The participants were 60 undergraduate students at University of California, Santa Barbara, who received either course credit or $8.00 for their participation. Participants were randomly assigned to one of three groups: 20 participated in the hypermedia manual condition, 20 participated in the full text condition, and 20 participated in the causal text condition.

Materials

Manual: One group of participants learned how the toilet tank worked by interacting with the hypermedia manual described above.

Full text study materials: The second group of participants received a printed version of all the information presented in the manual. This consisted of nine printed pages, each of which showed a fully labeled schematic diagram of the toilet tank. Text on the first page described the overall purpose of the toilet tank so that this page corresponded to section 1 of the manual. The second page text (labeled “subsections”) described the subsystems and corresponded to section two of the manual. The next three pages (labeled “connections”) described each of the components of the toilet tank in turn, presenting the text from section 3 of the manual. The following two pages (labeled “questions”) presented the multiple choice questions from section four of the manual. Then one page described the causal chain and movement of the toilet tank (corresponding to sections 5 and 6) and finally, the last page contained the explanation of the siphon from section 7 of the manual.

Causal text study materials: The third group of participants received a printed version of three sections of the manual that was printed on three pages. The first page showed the schematic labeled diagram of the toilet tank and described the overall function of the tank (corresponding to section 1 of the manual). The second page described the causal chain and movement of the toilet tank (corresponding to sections 5 and 6 of the manual). The third page contained the explanation of the siphon.

Test questions: The test questions consisted of 15 multiple choice questions and 11 open ended questions about the toilet tank. Ten of the questions, which we will refer to as “mental animation” questions required people to predict how a component of the system will be moving, given that another component is moving in a specified way. For example, one question asked “Imagine the connecting rod is moving up. What is happening to the float arm?”. There were five answer choices for each question, one which described the correct direction of motion, another that described the correct direction of motion but said that it would happen in the future rather than simultaneously (which was given half credit). For example the correct answer to the sample question above is “it is moving down”. Half credit would be given for the answer “It is not moving now, but it will move down soon”. The other three answer choices were incorrect. Five other multiple choice questions asked about principles such as the siphon effect and other principles underlying the operation of the toilet tank and will not be analyzed in this paper.

The short-answer questions were 11 open ended questions about the toilet tank. Three “function” questions asked the function of a component in the system, e.g., “What is the function of the float and float arm?”. Four “fault-behavior” questions that described a particular fault in a component of the system and asked participants to predict how the system would behave, e.g., “How would the tank function be affected if the inlet valve was stuck in the water inlet pipe? (list all possible answers)”. Two troubleshooting questions described faulty behavior of the system and asked what components might be faulty, e.g. “Suppose that after flushing the toilet, you notice that water is continuously running into the tank. What could be wrong? (list all possible answers)”

Background questionnaire: A background questionnaire asked participants to report their quantitative and verbal SAT scores, to list any courses they had taken in physics, mechanics or mechanical engineering, to list any mechanical or electrical items that they had attempted to explain and specifically whether they had ever tried to fix a toilet, change the oil in a car or unblock a drain. Finally they were asked to rate on a scale of 1 to 7 how interesting they thought the material was (with 1 meaning not interesting at all and 7 meaning very interesting).

Procedure

Participants were tested one at a time. When they arrived in the laboratory they were seated either at an empty table (in the full text and causal text conditions) or at a table containing a Power Macintosh 8500/180 with a 17 inch monitor (in the multimedia condition). Participants were instructed that they would be asked to read descriptions and view diagrams of a toilet tank and that they should try to understand the toilet tank so that they could explain to another person how it worked. They were informed that after studying the materials, they would be asked to answer questions testing their understanding of the toilet tank.

Participants were allowed as much time as they wished to study the materials and were told to let the experimenter know when they were ready to answer the questions. The time taken to study the materials was measured using a stopwatch. Then participants were given the question booklet. They answered the multiple choice questions first and were allowed 15 minutes to answer these questions. Then they were asked to turn over to the short answer questions and were allowed 15 minutes to answer these questions. Finally they completed the questionnaire.

Results

Study Times: There were large differences in study times between the three groups ($F(2, 57) = 47.11$, $p < .001$). The causal text group spent less time (4.62 minutes, $SD = 1.89$) than each of the other groups, which is perhaps not surprising, given that they received less information. Although the full text and hypermedia manual groups received the same text, the latter group spent longer (14.57 minutes, $SD = 4.09$) than the full text group (10.43 minutes, $SD = 3.39$). This reflects time interacting with the computer interface and may also reflect the fact that participants had to view the whole animations in sections 5 and 6 of the manual, which were played at a constant rate. This rate might have been slower than the time taken to read the relevant text and integrate it with the diagram in the full (i.e., printed) text conditions.

Mental Animation Questions: In scoring the mental animation questions, participants were given a score of 2 for each correct answer and a score of 1 if they described the correct motion of the component, but said that this motion would happen in the future, rather than simultaneously. The total possible score was 20. As shown in Figure 2, there was no significant difference between the groups receiving the three types of instruction ($F(2, 57) = 0.07$). Participants were 55.4% correct on these problems on average, indicating that this result is not a ceiling effect. Chance performance on these questions is 30% correct, indicating that it is not a floor effect either.

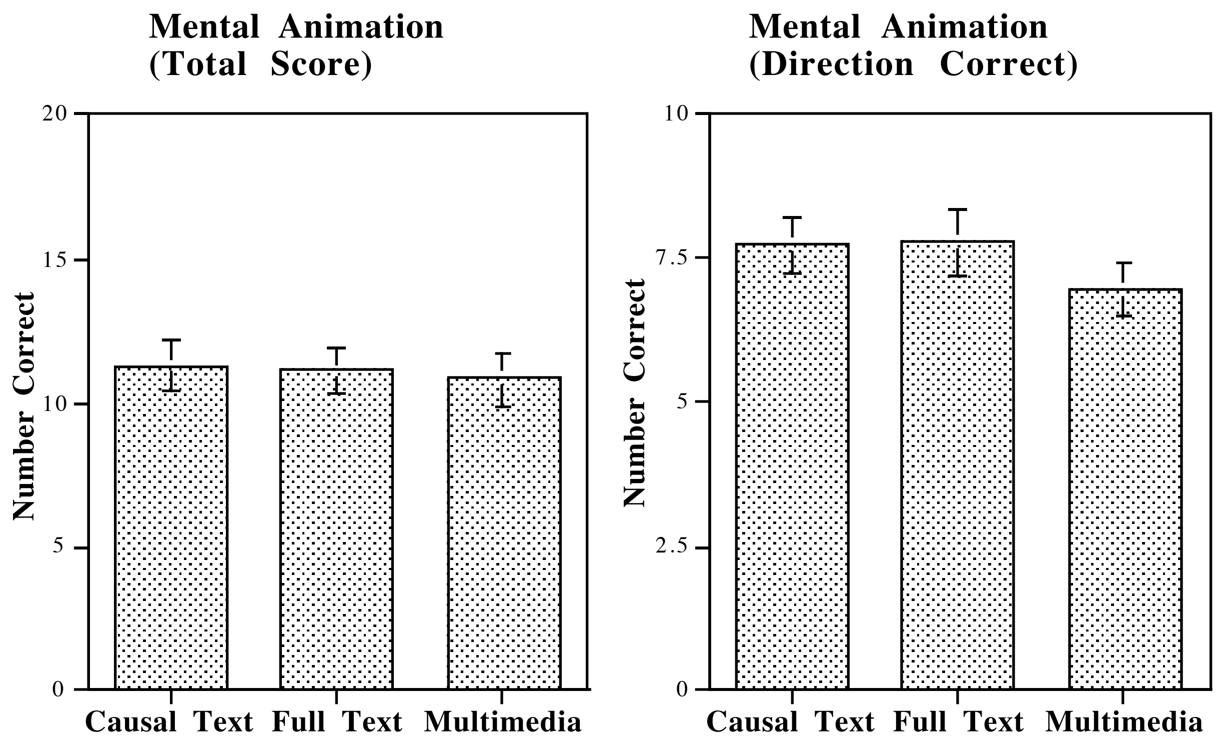

Figure 2: Performance on mental animation questions in Experiment 1. The first measure scores whether the participant answered the correct direction of motion and the correct timing of the motion. The second measure scores only the correct direction of motion. The maximum score on the y-axis reflects the maximum possible score. Error bars in this and all graphs show standard error of the mean.

In an alternative scoring of the ten mental animation questions, we scored an answer as correct (1 point) if the participant described the correct direction of motion for the component in question, regardless of whether they said the motion would occur simultaneously or "soon". Therefore the total possible score on this measure was 10. There was no significant difference between the groups on this measure ($F (2, 57) = 0.78$). On average, subjects were 75% correct on this measure compared to a chance level of 40% correct.

Prior knowledge and experience affected participants performance on the mental animation questions, but did not interact with the type of instruction. Thirty-one of the participants had taken at least one physics or engineering course and 29 had not. Physics training was associated with higher performance on the second measure of mental animation that takes into account whether only the correct direction of motion ($F(1, 54) = 6.07$; $p < .05$). On average, participants reported that they had repaired 3.43 items and we classified participants as having more practical experience if they had repaired 4 or more items and less practical experience if they had repaired less than 4 items. Participants with more practical experience had higher scores on the first measure of mental animation, that takes into account both the correct direction of motion and the timing of that motion ($F(1, 54) = 6.24, p < .05$). None of the measures of training or experience had significant interactions with the type of instruction suggesting that the different types of instruction were not differentially effective for individuals with different amounts of prior knowledge or experience.

Open Ended Questions: We analyzed scores on the three major types of open-ended questions -- the questions about the function of a specific component, fault-behavior questions and troubleshooting questions. These scores are shown in Figure 3. None of these measures differed as a function of type of instruction ($F(2, 57) < 1.1$ in all cases). Furthermore, neither physics training or practical experience affected these measures and there were no significant effects of the interaction of training or experience with the type of instruction.

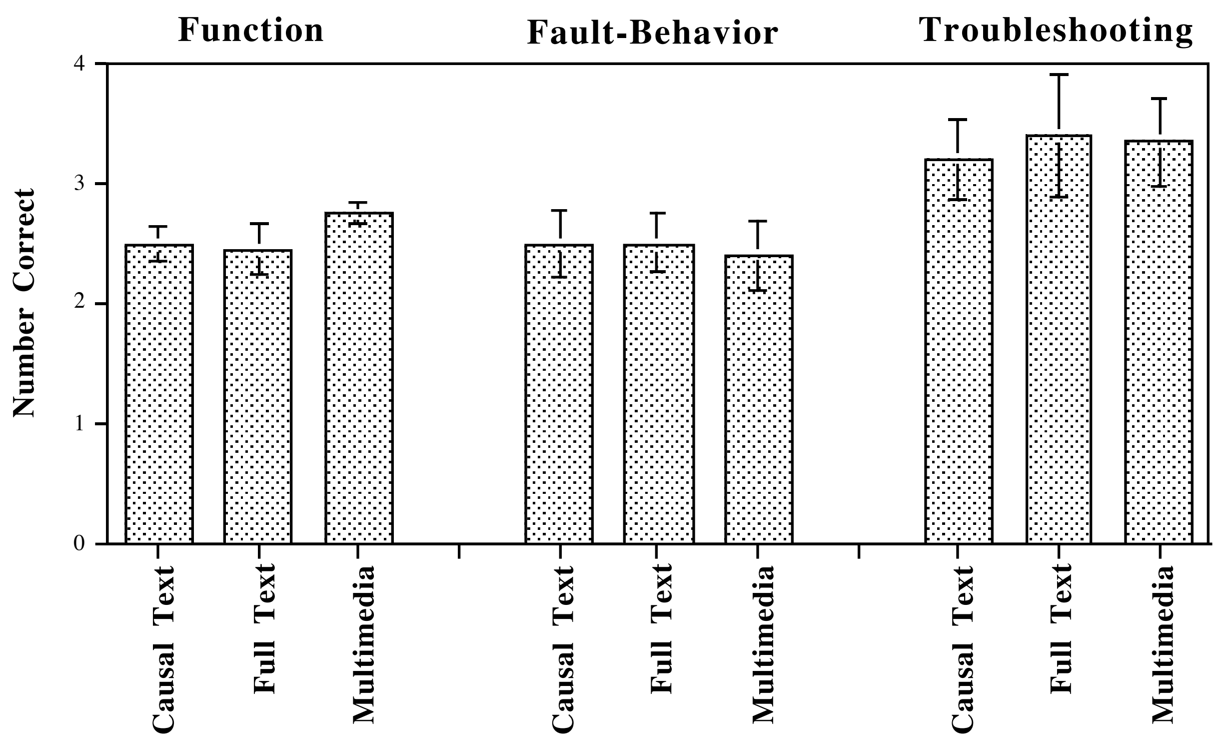

Figure 3: Performance on the open-ended questions in Experiment 1.

Interest Ratings: The type of instruction did not influence participants ratings of interest in the materials. The mean rating was 4.05 for the hypermedia group, 4.10 for the full text group and 4.30 for the causal text group ($F(2, 57) = 0.12$).

Discussion

Experiment 1 showed no differences in learning outcomes between the three groups. The similarity in performance between the full text and hypermedia groups indicates that whether people view information on a computer screen, interact with a hypermedia interface, receive text that is hyperlinked with diagrams, or view animations with commentaries rather than printed text and static diagrams does not have significant effects on learning.

A comparison of the full text and hypermedia groups with the causal text groups also reveals no differences in learning outcomes. This suggests that explicitly describing the functions of the toilet tank subsystems, the connections between components and the material composition of the components does not affect people's ability to understand the causal chain or kinematics of the device. On first glance this appears to be in contradiction to the principles of our theoretical model. However, recall that subjects in the causal text condition were first given a labeled diagram of the device. This diagram provided information about the decomposition of the diagram and the spatial relations between components. Therefore the additional text describing these components in the other conditions might have been superfluous. Furthermore, since a toilet tank is a household item, people might have knowledge of the functions and material composition of its parts as a result of background knowledge so that the additional information on these topics in the full text and multimedia conditions is superfluous.

It is possible that the equivalence of learning outcomes in Experiment 1 is specific to the particular machine that we chose to explain. Comprehension of this machine might not be sensitive to differences in instructional treatments either because participants in our study are already quite familiar with it or because particular causal and kinematic relations between components are too difficult to understand on the basis of the short instruction given here. Therefore it is important to consider whether the results of this experiment generalize to comprehension of other machines. In Experiment 2 we examined comprehension of three machines, the toilet tank, car brakes and a bicycle pumps. The latter two machines have been the topic of instruction in several previous studies of multimedia learning (Mayer, 1989) in which large differences in learning were found between different instructional conditions. This indicates that these machines are comprehensible with the right instruction by the population of study here (college students) and also that prior knowledge of these machines is not so great as to eliminate effects of instruction.

Experiment 2

Method

Participants: The participants were 66 undergraduate students who received either course credit or $10 for their participation. Participants were randomly assigned to one of three groups: 21 participated in the diagram only condition, 23 participated in the causal text condition, and 22 participated in the full text condition.

Materials

Study Materials: There were three sets of study materials for each of the three machines. The <u>full text</u> study materials for each machine consisted of two pages, each of which included a labeled diagram of the machine. On the first page this was accompanied by a text describing the material composition of each of the components of the machine and their spatial relations (connections). On the second page, this was accompanied by a text that described the causal sequence of motions that results when the machine is operated (i.e. the handle on the toilet tank is pushed down, the driver steps on the car's brake pedal and a user pushes down and pulls up the handle on a bicycle pump. The <u>causal</u> text materials consisted of only the second page of the full text materials. The <u>diagram-only</u> study materials consisted of just the labeled diagrams of each of the three machines.

Test questions: The test questions consisted of 18 multiple choice questions about the toilet tank, 16 multiple choice questions about the bicycle tire pump, and 12 multiple choice questions about the car brakes in the form, "Imagine machine component A is changing in some particular way, what is happening to machine component B?" The questions varied the number of steps in the causal chain that had to be inferred, and depended on the length of the causal chain for each mechanical system. For the toilet tank question one question involved inferring 7 steps in the causal chain, two questions involved inferring 2 steps in the causal chain and 3 questions each involved inferring 1 to 5 steps. For the brakes questions one question involved inferring 5 steps, two involved inferring 4 steps and three questions each involved inferring 1 to 3 steps. For the pump questions, two questions involved inferring 6 steps in the causal chain, two involved inferring 5 steps and three questions each involved inferring 1 to 4 steps.

In addition, participants solved 4 problem solving questions for each of the machines, that included troubleshooting questions, and questions about how the machines could be made more reliable and more effective. The problem solving questions for the brakes and pumps had been used in previous studies (Mayer, 1989).

Background questionnaire: The background questionnaire was identical to that used in Experiment 1 except that it also included specific questions asking participants whether they had ever used a tire pump or worked on car brakes and it did not include the question about how interesting participants found the materials.

Procedure

Participants were tested in groups of 1 to 5. Each group of participants was assigned to either the full text condition, the causal text condition or the diagram only condition. The order in which they read about the three different machines (toilet tank, car brakes and bicycle pump) was counterbalanced across groups so that approximately equal numbers of subjects studied the machines in each of the six possible orderings.

Participants were given the same general instructions as in Experiment 1. Then they were given the study materials for the first machine and were allowed a specified amount of time to study these materials -- 7 minutes for the toilet tank, 4 minutes for the car brakes and 4 minutes for the bicycle pump. After studying each machine, participants answered the questions about that machine. They were first given a specified amount of time to answer the multiple choice (mental animation) questions for the machine, which was roughly proportional to the number of questions to be answered -- 6 minutes for the toilet tank questions, 4 minutes for the car brakes and 5 minutes for the bicycle pump. After answering the multiple choice questions for a machine, participants were then given 2 minutes to answer each of the short answer questions about that device. This procedure was repeated for the second and third machines in the ordering.

Finally participants answered the background questionnaire, and were thanked and dismissed.

Results

Multiple Choice (Mental Animation) Questions: As in Experiment 1, participants were given a score of 2 for each correct answer and a score of 1 if they described the correct motion of the component, but said that this motion would happen in the future, rather than simultaneously. We computed total scores for the pumps, brakes and toilet tank questions respectively and these are shown in Figure 4. The total possible score was 36 for the toilet tank questions, 24 for the brakes questions and 32 for the bicycle pump questions. Data were analyzed separately for each of the three machines.

As shown in Figure 1, there was no effect of instruction type for the toilet tank questions (F (2, 63) = 1.144, p = 32) or the brakes questions (F (2, 63) = .541, p=.58). There was a significant difference between the conditions in the case of the bicycle pump questions (F (2, 63) = 5.16, p < .01. Participants in the causal text and full text conditions were more accurate than subjects in the diagram only condition, but did not differ from each other.

As in Experiment 1, prior knowledge and experience affected participants' performance on the mental animation questions, but did not interact with the type of instruction. Exactly half of the participants in the experiment had taken at least one physics or engineering course and half had not.

Physics training was associated with higher performance on the brakes questions ($F(1, 60) = 6.22$, $p < .05$) and had a marginal effect on performance on the pumps questions ($F(1, 60) = 2.66$, $p = .11$) but no significant effect on the toilet tank questions.

In terms of practical experience, 59 participants had used a tire pump and 7 had not, 12 participants had tried to work on car brakes and 54 had not and 47 had tried to fix a toilet and 19 had not. On average, participants reported that they had repaired 5.08 items and we classified participants as having more experience if they had repaired 6 or more items and less practical experience if they had repaired less than 6 items. Participants with more practical experience had higher scores on the brakes questions ($F(1, 60) = 5.11$, $p<.05$) and there was a marginal effect of practical experience on scores on the pumps questions ($F(1, 60) = 3.49$, $p = .07$), but no effect on the toilet tank questions. None of the measures of training or experience had significant interactions with the type of instruction indicating that the instruction conditions were not differentially effective for individuals with different amounts of prior knowledge or experience.

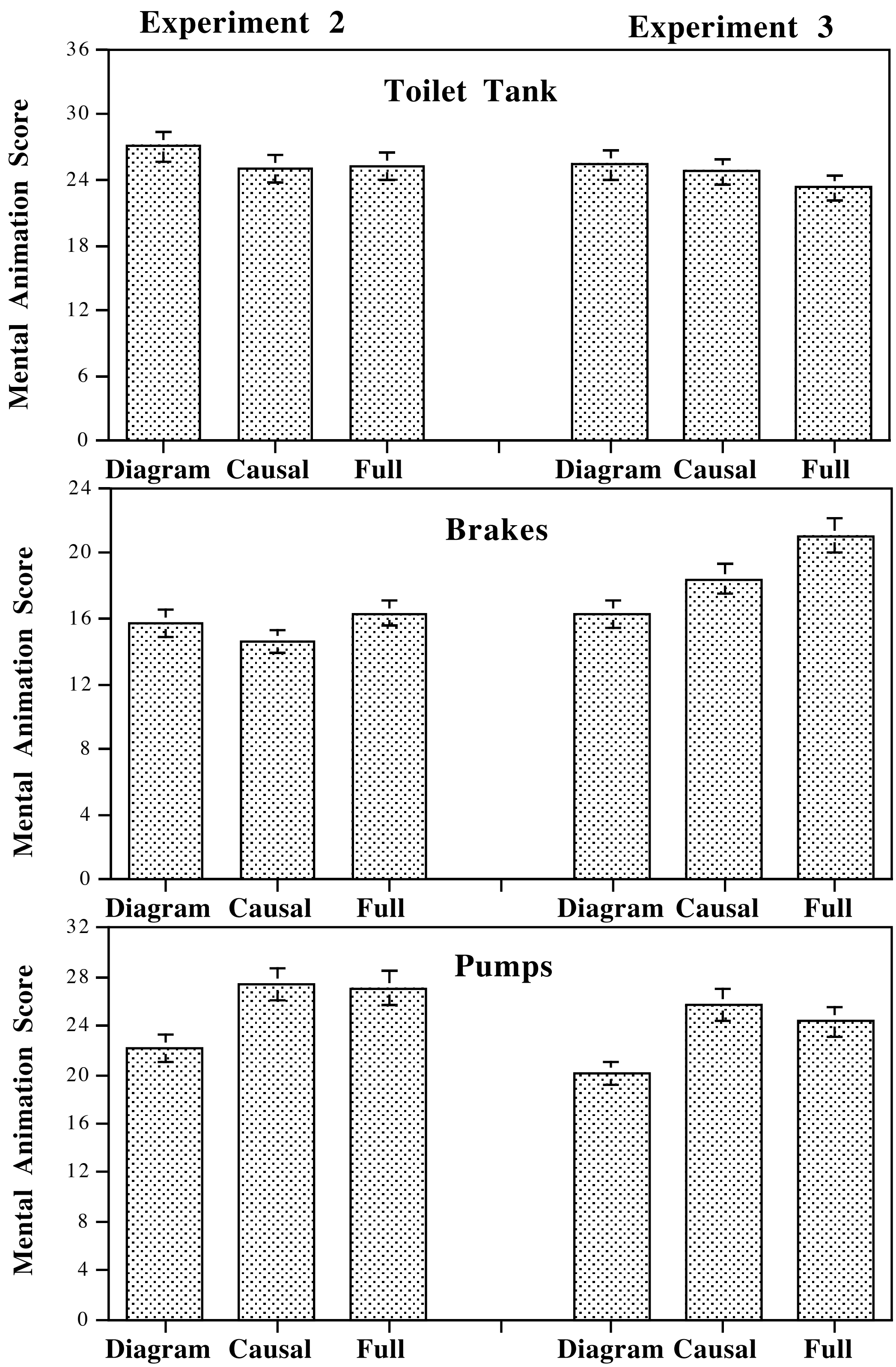

Figure 4: Mental Animation Scores for Experiments 2 and 3.

Problem Solving Questions: Two independent raters scored the 4 problem solving questions for each of the three machines and we computed a composite score for each machine by adding the scores for the 4 questions about that machine. The inter-rater reliability was .85 for the pump questions, .95 for the brakes questions and .85 for the toilet tank questions. Problem solving performance is graphed in Figure 5.

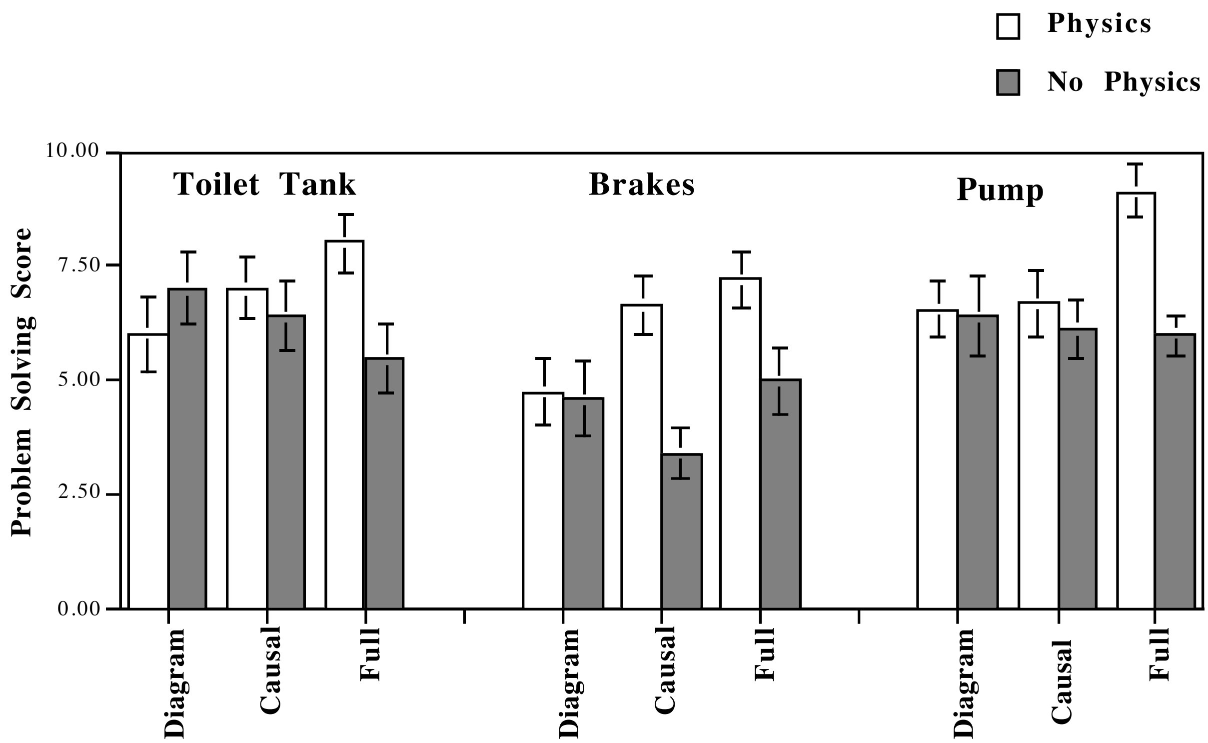

Figure 5: Performance on the problem solving questions in Experiment 2. Data is presented separately for the three mechanical systems and for participants who had and had not ever taken a physics course.

We first compared performance of the three instructional groups on the problem solving questions. Overall there were no significant effects for any of the machines ($F\ (2, 63) = .064$, $p = .94$ for the toilet tank questions; $F\ (2, 63) = 1.307$, $p = .28$ for the brakes questions; $F\ (2, 63) = .947$, $p = .39$ for the pump questions).

For all three of the machines, there was a trend for participants with physics training to benefit more from text accompanying the diagrams. A marginal effect of the interaction between physics training and type of instruction on the toilet tank questions ($F\ (2, 60) = 2.93$, $p = .06$) indicated that addition of the text tended to help participants who had taken a physics course but to hinder participants who had not taken physics. Participants with physics training had higher scores on the brakes questions ($F\ (1, 60) = 10.52$, $p < .01$) and there was again a marginal interaction of training with instruction type ($F\ (2, 60) = 2.53$, $p = .09$) suggesting that text tended to help these participants more. Participants with physics training also had superior performance on the pumps questions ($F\ (1, 60) = 5.84$, $p = .02$) and there was again a marginal interaction with type of instruction ($F(2, 60) = 2.99$, $p = .06$). This result indicated that for this machine, the full text but not the causal text tended to help participants with physics training.

In contrast, there was no interaction of instruction with amount of practical experience with machines. Participants who had more practical experience with machines were better able to solve problems about the toilet tank ($F\ (1, 60) = 3.76$, $p = .06$) and bicycle pump ($F(1, 60) = 5.29$, $p <$

.05) but not car brakes. The interaction of instruction with practical experience did not affect problem solving questions for any of three machines.

Discussion

This experiment replicated Experiment 1 in that it showed that people who studied the full text describing the toilet tank were no more able to mentally animate the toilet tank than people who studied the causal text. Furthermore this result generalized to comprehension of two other machines -- car brakes and a bicycle pump.

An even more striking result of this experiment was that there were no advantages of diagrams accompanied by text (either full or causal) over labeled diagrams alone for mental animation of either the toilet tank or the car brakes. Only in the case of the bicycle pump was there an advantage of an accompanying text. In the case of this machine, the text describing the causal chain and movement described the pressure changes in the pump cylinder that result from operating the pump. Since pressure is not a visible property, it might not have been possible to infer information about pressure from the diagram alone. In contrast, the text describing the causal chain and movement of the toilet tank and car brakes described only the movement of visible components of the machines. This experiment suggests that such information can be inferred from the diagram without an explanatory text.

Performance on the problem solving questions showed more sensitivity to the different forms of instruction at least in the case of participants who had some prior knowledge of physics. For all three machines, there was a trend for participants who had some prior training in physics to benefit more from text accompanying the labeled diagrams. There was no such trend for participants who had not taken physics. This result suggests that some knowledge of physics principles might be a prerequisite for learning from the texts presented here.

Experiment 3

It is possible that the lack of instructional effects in Experiments 1 and 2 were due to the type of assessment used. In particular our main measure of how well subjects could run a mental model of the machines was based on multiple choice questions. It is possible that these types of questions are not maximally discriminating between the conditions because some of the answer choices can easily be eliminated on the basis of background knowledge. Therefore in Experiment 3 we attempted to replicate Experiment 2 but with mental animation questions that were open ended rather than multiple choice.

Method

Participants: The participants were 60 undergraduate students who received either course credit or $10 for their participation. Participants were randomly assigned to one of three groups: 18 participated in the diagram only condition, 23 participated in the causal text condition, and 19 participated in the full text condition.

Materials

The instruction sheet, and study materials (full text and diagram, causal text and diagram and diagram only) were identical to those used in Experiment 2.

The test questions consisted of open-ended questions that corresponded to the multiple choice questions used in Experiment 2. Again there were 18 questions about the toilet tank, 16 questions about the car brakes and 12 questions about the bicycle pump. These questions asked about the same specific components of the machines but required an open-ended response and were phrased as follows: "Imagine machine component A is changing in some particular way. What is happening to machine component B?"

The background questionnaire was identical to that used in Experiment 2.

Procedure

As in Experiment 2, participants were tested in groups of 1 to 5 and each group was assigned either the full text, causal text, or diagram only condition and to a specific ordering of the three machines, such that approximately equal numbers of subjects studied the machines in each of the six possible orderings.

The time allotted to study each of the machines was the same as in Experiment 2. After studying each machine, subjects were given a specified amount of time to answer the open-ended questions about that machine -- 9 minutes for the toilet-tank questions, 6 minutes for the brakes questions and 8 minutes for the bicycle pump questions. Finally participants answered the background questionnaire, and were thanked and dismissed.

Results

As in Experiment 2, subjects were given 2 points for a question if described the correct motion of the component in question and one point if they described the correct motion but said it would happen in the future rather than simultaneously. The results are presented in Figure 4. As in Experiment 2, instruction had no significant effect on answers to the toilet tank questions ($F(2, 57) = 0.319$, $p=.73$). In this experiment there was a marginal effect of instruction on the brakes questions ($F(2, 57) = 2.549$, $p=.09$) such that participants in the full text condition had the highest scores and participants in the diagram condition had the lowest scores. Finally, the significant effect of text accompanying the diagrams for the pumps questions, found in Experiment 2, was replicated here ($F(2, 57) = 4.67$; $p = .01$).

Thirty-nine of the participants in the experiment had taken at least one physics or engineering course and 21 had not. Physics training was associated with higher performance on the toilet tank questions ($F(1, 54) = 3.89$, $p = .06$) but there was no significant effect of training on mental animation of the other machines. In terms of practical experience, 58 participants had used a bicycle pump and 2 had not, 9 participants had tried to work on car brakes and 51 had not and 31 had tried to fix a toilet and 9 had not. On average, participants reported that they had repaired 4.12 items and we classified participants as having more practical experience if they had repaired 4 or more items and less practical experience if they had repaired less than 4 items. Practical experience had no effect on the mental animation questions. As in previous experiments, none of the measures of training or experience had significant interactions with the type of instruction.

Discussion

In general , the level of performance on the mental animation questions in this experiment was similar to that in Experiment 2, indicating that the failure to find effects of instruction in Experiments 1 and 2 is not primarily due to the assessment method. As in Experiment 2, instruction had no effect on ability to mentally animate the toilet tank and the two text groups were better able to mentally animate the bicycle pump. However there was a trend suggesting that additional text facilitated ability to mentally animate the car brakes, which was not observed in Experiment 2, suggesting that at least for this mechanical system, there might be an effect of instruction.

5. General Discussion

In summary, we developed a model of how people understand mechanical systems from text and diagrams (Narayanan & Hegarty, in press). This model made two types of recommendations for how to develop instructional manuals that explain how machines work. One type of recommendation concerned the content of the manual, (information about the decomposition of the diagram, connectivity between the components etc.). This recommendation is applicable to both hypermedia and more traditional printed manuals. The other type of recommendation concerned how hypermedia capabilities (e.g., hyperlinks and animation) can be used to facilitate comprehension.

Experiment 1 suggested that providing additional information about the decomposition of the machine, connectivity between the components and material composition of the components had no effect on comprehension of a relatively complex mechanical system -- a toilet tank. This result was replicated in Experiments 2 and 3. Furthermore, Experiment 1 also showed that presenting information in a hypermedia manual that includes hyperlinks, colored diagrams, animations (rather than static diagrams) and commentaries (rather than visual text) had no effects on comprehension.

Experiments 2 and 3 generalized the first result to comprehension of other machines -- car brakes and a bicycle pump. In none of the experiments did studying the more complete text lead to significantly better comprehension compared to text that merely described the causal chain and movement of the system. Furthermore, Experiments 2 and 3 went on to show that in some cases, viewing a labeled diagram alone lead to the same level of comprehension as viewing diagrams accompanied by text. Addition of the text improved understanding of the bicycle pump in both experiments and the car brakes in one experiment, but had no effect on comprehension of the toilet tank.

Do these results suggest that we should reject our model of comprehension and corresponding design guidelines? We believe that it would be premature to reject the model at this stage. For example, it is possible that the machines that participants studied in these experiments were already familiar to them, so that information about their decomposition and function was already known. Furthermore, the configuration of these machines might have been easily comprehended from a static diagram, so that additional text describing the configuration of components might have been superfluous. It is also possible that our outcome measures were not sensitive to differences in comprehension resulting from the different instructional methods. Our experiments showed some evidence that more open ended comprehension questions such as the problem solving questions used in Experiment 2 and the mental animation questions used in Experiment 3 might be more sensitive to the instructional treatments.

In these experiments the multimedia (text and diagram) conditions were designed to provide participants with guidance in interpreting the diagrams. However, these experiments suggest that such guidance might not be necessary. Participants were able to make some inferences about how relatively complex machines worked from a diagram alone. These results suggest that rather than the highly directive nature of the instructional manuals provided here, a better application of hypermedia might be a system in which users are *encouraged to make inferences* from diagrams *to construct* knowledge about how mechanical systems work. More intelligent hypermedia manuals that can assess a user's prior knowledge and focus on presenting information that he or she does not already know, might also be more effective in enabling people to construct accurate mental models. Finally, these experiments provide some indication that the general belief in the instructional benefits of animations needs to be revisited. When mental animations (Hegarty, 1992) are feasible from static multimodal presentations, benefits of system-provided animations may not be sufficient to warrant the costs of building such animations.

In future experiments we plan to compare dynamic (hypermedia) and static (text and diagrams) multimodal information presentations designed according to our theoretical model against more traditionally designed dynamic (e.g., the "How Things Work" CD-ROM) and static (e.g., printed manuals) presentations. Another research avenue is to compare the guided navigational model we have implemented in the present prototype against one that allows fully unconstrained inter- and intra-section navigation. A third direction is to track subject eye fixations as they study parts of our multimedia system to study the on-line comprehension processes. This method has been used productively by Faraday & Sutcliffe (1997a, 1997b) as a basis for deriving design guidelines for multimedia presentations.

It is instructive to compare our results to those of Faraday & Sutcliffe (1997a; 1997b). They took a different approach from ours in that they derived guidelines for multimedia design from experiments with an existing multimedia presentation system. They found that redesign based on empirically derived design guidelines not only improved recall, but also that images and animations contributed to this improvement There are a number of differences between their studies and ours. First, their domain was cellular processes whereas ours involved a device of everyday use. Second, they tested recall whereas we measured deeper comprehension by asking questions that require subjects to make new inferences from the presented material. It is possible that animations are effective for conveying information about unfamiliar or abstract processes, but are not necessary when users know enough to engage in mental animation. It is also possible that animations contribute to better recall of the presented information, but not to deeper comprehension that facilitates inferring new information.

Similarly, studies by Mayer and Sims (1994) have shown the benefits of spatially and temporally synchronized multimedia presentations over asynchronous ones. However they have not directly compared the benefits of dynamic presentations such as narrated animations to those from static mixed-mode presentations. Therefore, our results are complementary, not contrary to Mayer & Sims' results. Clearly, these strands of research have unearthed a number of open issues regarding the necessity and benefits of the various features of hypermedia information presentations.

References

Biederman, I. (1987). Recognition-by-components. A theory of human image understanding. *Psychological Review*, 94, 115-147.

Chi, M. T. H., de Leeuw, N., Chiu, M. & LaVancher, C. (1994). Eliciting self-explanations improves learning. *Cognitive Science*, 18, 439-478.

Faraday, P. & Sutcliffe, A. (1997a). An empirical study of attending and comprehending multimedia presentations. *Proc. ACM Multimedia 96 Conference,* 265-275, ACM Press.

Faraday, P. & Sutcliffe, A. (1997b). Designing effective multimedia presentations. *Proc. ACM Conference on Human Factors in Computing Systems (CHI'97),* 272-278, ACM Press.

Hegarty, M. (1992). Mental animation: Inferring motion from static diagrams of mechanical systems. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 18(5), 1084-1102.

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

Marr, D. & Nishihara, H. K. (1978). Representation and recognition of the spatial organization of three-dimensional shapes. *Proceedings of the Royal Society*, Vol B 200, 269-294.

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

Mayer, R. E. & Sims, V. K. (1994). For whom is a picture worth a thousand words? Extensions of a dual-coding theory of multimedia learning. *Journal of Educational Psychology*, 86, 389-401.

Narayanan, N. H. & Hegarty, M. (in press). On designing comprehensible interactive hypermedia manuals. *International Journal of Human-Computer Studies*, *48*.

Narayanan, N. H., Suwa, M. & Motoda, H. (1994a). How things appear to work: Predicting behaviors from machine diagrams. *Proceedings of the 12th National Conference on Artificial Intelligence*, AAAI Press, 1161-1167.

Narayanan, N. H., Suwa, M. & Motoda, H. (1994b). A study of diagrammatic reasoning from verbal and gestural data. *Proceedings of the 16th Annual Conference of the Cognitive Science Society*, Lawrence Erlbaum Associates, 652-657.

Novak, G. (1995). Diagrams for solving physics problems. In J. I. Glasgow, N. H. Narayanan & B. Chandrasekaran. (Eds.). *Diagrammatic Reasoning: Computational and Cognitive Perspectives*, Boston, MA: MIT Press and Menlo Park, CA: AAAI Press.

Scaife, M. & Rogers, Y. (1996) External cognition: How do graphical representations work. *International journal of Human-Computer Studies, 45,* 185-213.

Schwartz, D. L. & Black, J. B. (1996). Analog imagery in mental model reasoning: Depictive models. *Cognitive Psychology*.

Author Notes:

Mary Hegarty and Jill Quilici, Department of Psychology, University of California, Santa Barbara, CA 93106. N. Hari Narayanan and Selma Holmquist, Department of Computer Science and Engineering, Auburn University, Auburn AL 36849.

The authors acknowledge financial support for this research from the Office of Naval Research through contract N00014-96-1-0525 to Mary Hegarty and contract N00014-96-1-1187 to Hari Narayanan.