Available online at www.sciencedirect.com
ScienceDirect
ELSEVIER
Int. J. Human-Computer Studies 65 (2007) 911–930
International Journal of Human-Computer Studies
www.elsevier.com/locate/ijhcs

Top-down and bottom-up influences on learning from animations

Sarah Kriz, Mary Hegarty*

Department of Psychology, University of California, Santa Barbara, CA 93106-9660, USA

Received 22 November 2006; received in revised form 22 June 2007; accepted 26 June 2007
Communicated by D. Boehm-Davis
Available online 1 July 2007

Abstract

To evaluate how top-down and bottom-up processes contribute to learning from animated displays, we conducted four experiments that varied either in the design of animations or the prior knowledge of the learners. Experiments 1–3 examined whether adding interactivity and signaling to an animation benefits learners in developing a mental model of a mechanical system. Although learners utilized interactive controls and signaling devices, their comprehension of the system was no better than that of learners who saw animations without these design features. Furthermore, the majority of participants developed a mental model of the system that was incorrect and inconsistent with information displayed in the animation. Experiment 4 tested effects of domain knowledge and found, surprisingly, that even some learners with high domain knowledge initially constructed the incorrect mental model. After multiple exposures to the materials, the high knowledge learners revised their mental models to the correct one, while the low-knowledge learners maintained their erroneous models. These results suggest that learning from animations involves a complex interplay between top-down and bottom-up processes and that more emphasis should be placed on how prior knowledge is applied to interpreting animations.

Keywords: Animations; Mental Models; Interactivity; Learning

1. Introduction

Throughout history, external visualizations (e.g., diagrams, graphs, maps, and schematic drawings) have been produced by inventors, scientists, and navigators to convey scientific and technical information (Ferguson, 1992). One area in which external visualizations are especially valuable is in the domain of mechanics. Because machines are spatially and causally complex, many aspects of the structure and kinematics of mechanical systems can be more effectively conveyed by external visual-spatial representations than by verbal descriptions (cf. Larkin and Simon, 1987). For instance, an assembly diagram of an electric garage door opener can be more effective in illustrating the parts and workings of the mechanism than a purely textual explanation containing the same information.

As personal computers have become more accessible, educators have turned to electronic media in order to teach students about how physical systems work (e.g., Macaulay, 1998). This shift from printed to electronic media has enabled instructors to convey information using animations rather than static visualizations. At first glance, there appear to be many practical benefits to using animations. Animations can directly and explicitly represent the movements of mechanical parts (i.e., the external representation of behavior can be isomorphic to reality). Whereas the actual movements of a machine must be viewed in real time, an animation can display movements that are either too fast or too slow to perceive in real time.

Although animations can present information about movement in a more explicit way, the benefits of learning from animations compared to static diagrams remain uncertain. Previous studies have found that when animations and static diagrams convey "equivalent" information there is no benefit of animations over static diagrams and text (e.g., Pane et al., 1996; Tversky et al., 2002; Hegarty et al., 2003). Furthermore, the temporal constraints of animations may increase "cognitive load" and actually lower performance when compared to learning from static

---

*Corresponding author. Tel.: +1 805 893 3750; fax: +1 805 893 4303.
E-mail address: hegarty@psych.ucsb.edu (M. Hegarty).

1071-5819/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ijhcs.2007.06.005

media (Mayer et al., 2005). However, a number of researchers have suggested that certain design features may be employed to avoid such problems. These include interactivity (ChanLin, 1998; Tversky, et al., 2002; Hegarty, 2004; Schwan and Riempp, 2004) and attentional cuing or signaling (Faraday and Sutcliffe, 1997; Grant and Spivey, 2003; Mayer and Moreno, 2003). By focusing on improving the ways in which information is delivered, these proposals emphasize a bottom-up theory of comprehension, that what is learned is primarily a function of the quality of the information display. Animation researchers have devoted less attention to top-down influences on comprehension, that is, the ways in which prior knowledge and expectations affect how people attend to the display and how the new information in the display is integrated with prior knowledge (cf. Kintsch, 1988).

The present study evaluates the effectiveness of animations from two perspectives. On the one hand, we examine how designing animations in accord with cognitive design principles affects learning outcomes, in essence, studying bottom-up influences on learning from animations. On the other hand, we examine the role prior knowledge plays in the interpretation of dynamic external visualizations, or, the effect of top-down processing. By evaluating the effects of both animation design and prior knowledge, we attempt to explain how people understand a complex system as a result of viewing animations, and the criteria necessary for successful learning.

As with any comprehension task, the goal of understanding an animation is to create an internal mental model that accurately represents the referents of the animation (i.e., the objects or events that the animation depicts). Fig. 1 presents an overview of the comprehension process when someone constructs a mental model from a visual display (adapted from Hegarty, 2005). One aspect of this process involves accurately encoding the information in the external display using bottom-up processes (shown by solid arrows in Fig. 1). This leads to an internal representation of the information explicitly conveyed in the display. Design features of the external visualization may affect what is encoded by displaying information more or less effectively. However, as highlighted by theories of text comprehension (Graesser et al., 1994; Kintsch, 1988) complex comprehension tasks also require that prior knowledge is integrated with sensory information in order to build a representation that is meaningful to the learner. Domain knowledge may influence the manner in which attention is directed to the external display and how information is perceived, interpreted, and modeled internally (these possible top-down processes are shown by the dashed arrows in Fig. 1). Thus, both the external representation and prior knowledge may have effects on how a mental model is constructed. Note that the overview model depicted in Fig. 1 is intended to be a general model of comprehension of any external representation, which might include text, diagrams, animations, etc. (Narayanan and Hegarty, 1998) but this paper evaluates the case of learning from animation without verbal instruction.

1.1. Bottom-up processing: the design of animations

Recent studies have suggested several reasons why animations may not be more effective than static diagrams, including their transience (Tversky et al., 2002; Ainsworth and Van Labeke, 2004), their demands on visual attention (Lowe, 1999; Mayer and Moreno, 2003), the fact that they encourage passive rather than active learning (Hegarty et al., 2003), and the fact that there can be a mismatch between the format of an animation and that of one's internal representation of motion (Tversky et al., 2002). In the present research, we focus on two of these issues: transience and demands on visual attention. With respect to the issue of transience, animations present a problem in that perception and comprehension processes must keep up with the pace at which the animation is presented, and information presented earlier in an animation must be

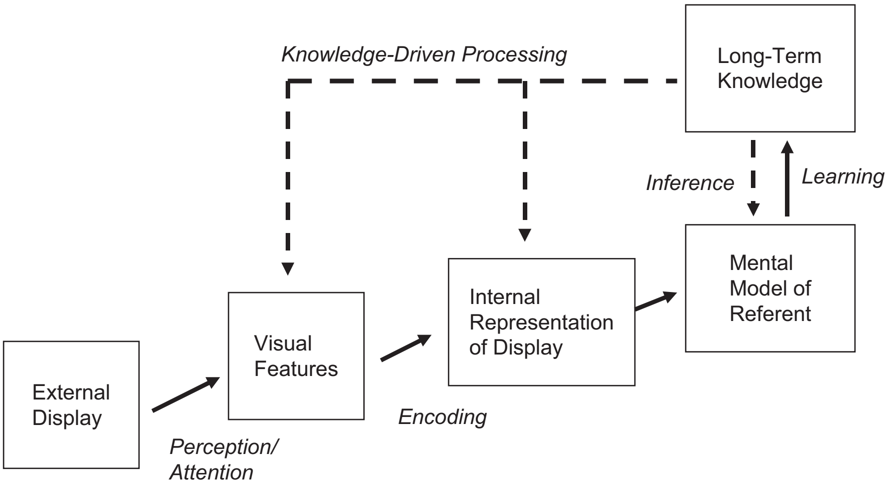

Fig. 1. Model of comprehension of visual displays, indicating the interaction of bottom-up processes (solid arrows) and top down processes (dashed arrows) in construction of a mental model of the referent.

maintained in working memory if it is to be integrated with information that is presented later (Lowe, 1999). When a learner views an animation, one frame is displayed at a time, and once the animation has advanced beyond a given frame, that frame is no longer available to the viewer.

Tversky et al. (2002) proposed that the problem of transience could be alleviated by making animations interactive. When an animation is interactive, users can slow down or speed up the animation to match their rate of perception and comprehension, and can return to previously viewed parts of the animation to help them integrate information presented at different points in time. In support of this theoretical claim, Chanlin (1998) found that students who received participant-controlled visual presentations performed better than students who used a system-imposed design (see also Schwan and Riempp, 2004). Other studies, however, have pointed to limitations in learning from interactive animations. For example, Pane et al. (1996) found no advantage of an interactive animation over an informationally equivalent computer presentation of text and still images. One possible reason for this result was that students did not take full advantage of the interactive controls, as found in more recent research by Lowe (2004). Thus, previously reported limitations on learning from interactive animations may stem from the lack of use of the features, rather than the ineffectiveness of the features themselves.

A second limiting factor of animations is the demands that they make on visual attention (Faraday and Sutcliffe, 1997; Mayer and Moreno, 2003). Animations of how machines work typically show several mechanical components moving at once. However, visual attention is spatially limited. There is a possibility that users of an animation may not attend to important events displayed in an animation because their attention is directed elsewhere. It is likely that learners have trouble attending to events in animations when the most salient motions are not those that are most thematically relevant (Lowe, 1999, 2003). Signaling or highlighting aspects of the display to make the most relevant information visually salient has been found to alleviate this problem somewhat (Faraday and Sutcliffe, 1997; Mayer and Moreno, 2003). Further evidence to support the effectiveness of signaling comes from the work of Grant and Spivey (2003), who found that participants who were trying to solve an insight problem (Dunker's Tumor problem) were more successful in arriving at the correct answer when they viewed a static diagram that visually highlighted the critical component of the diagram. This direction of visual attention appeared to lead participants to the correct answer, presumably because they began focusing on the area relevant for the solution of the problem. Thus, visual cueing can have an effect on how people construct understanding of external visual materials.

1.2. *Top-down processing: the effect of prior knowledge*

In addition to the design of an animation, a learner's prior knowledge also influences his or her success in learning from multimedia displays. Constructivist theories of comprehension (e.g., Kintsch, 1988; Chi et al., 1994; Graesser et al., 1994) assume that learning involves the integration of new information into existing knowledge structures. The result of this integration process depends not only on how the new information is presented, but also on the quantity, specificity, and accuracy of the existing knowledge. Specifically, prior domain knowledge may help a learner recognize what aspects he or she does not understand when acquiring new information. Miyake (1986) has proposed that developing a mental model of a mechanical system is often an iterative process, in which learners move back and forth between states of misunderstanding and understanding as they gain deeper knowledge of how the system works. In this process, Chi (2000) has suggested that revising a mental model can only occur when a conflict between the external representation and the internal representation is detected. Learners with a high level of prior domain knowledge should be better able to assess these "gaps" in their internal models or conflicts between internal and external representations. Consistent with this view, Rozenblit and Keil (2002) found that low domain knowledge (LDK) learners are often overconfident in judging their understanding of mechanical phenomena. Because of this overconfidence, "novice scientists" may not perceive any gaps or inconsistencies between their internal models and external animations.

Knowledge may also affect how a learner directs his or her visual attention while viewing a visual display such as an animation. While students with high domain-specific knowledge (e.g., physics knowledge) should attend to and integrate visual information on the basis of their pre-existing schemas, learners with only domain general knowledge (e.g., adults who have experienced physics principles in everyday life but have no formal physics training) may misdirect their visual attention to features of the display that are highly salient but not relevant. In fact, Lowe's (1994, 1999, 2003) studies in the domain of meteorology indicate that novices attend to perceptually salient rather than thematically relevant aspects of the displays. In other words, their processing of the animations is data driven rather than guided by prior knowledge.

1.3. *Overview of the study*

In this study, we examine how the design of instructional animations and prior knowledge affect people's mental models of a complex mechanical system, a flushing cistern. This specific type of mechanism is often used to flush toilets, but it differs from the mechanisms used in most American toilets (Fig. 2). While the purpose of the system is the same as an American model, the mechanism used to accomplish this function is different, so we assumed that

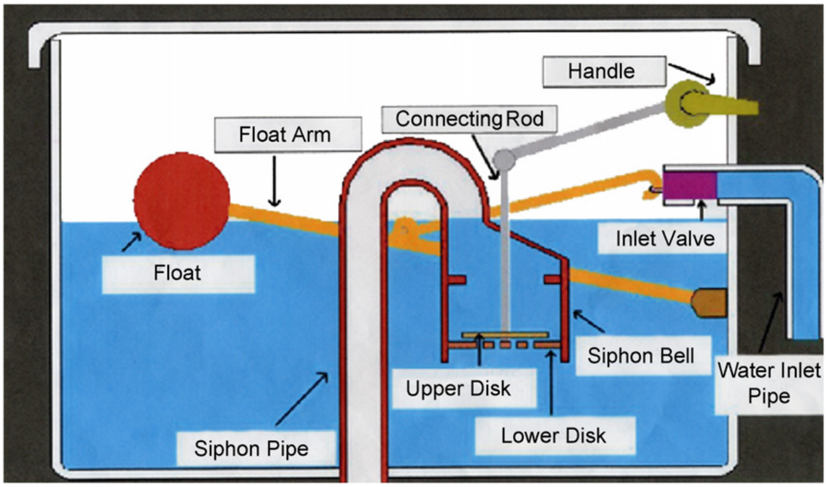

Fig. 2. Labeled diagram of the flushing cistern (toilet tank).

the participants in our experiments (American university students) had not previously seen or fixed this particular mechanical system.

The animations employed were relatively realistic. Realistic animations depict the system as it looks in the real world, including how and where the mechanical parts move (i.e., the kinematics of a mechanical system), but they do not depict invisible forces such as suction and buoyancy that do not have any direct visual manifestations. We do not claim that this is necessarily the best way to instruct people about how a machine works, nor was it our goal to design or test an optimal animation in this study. Recent research indicates that many people strongly believe that realistic displays, including real-time animations, are highly effective means of conveying complex visual information (Smallman and St. John, 2005). The goal of this study was to evaluate what people can and cannot actually learn from realistic animations, when given a general learning goal, so that we can better understand how we might augment or improve on these animations or on instructional activities with these animations in the future.

In Experiments 1–3, we assess how the design of realistic animations affects people's mental models of a flushing cistern, that is, how the quality of an external representation can affect the internal representation that a learner builds. In these experiments, students who had no formal education in physics or mechanics viewed animations, which varied in either interactivity or the presence of attentional signals, and were then asked to answer a series of questions about the flushing cistern. Experiment 4 assesses the contribution of prior knowledge. In this experiment students with high and low physics knowledge were given identical visual materials and asked to understand how the flushing cistern works. In this experiment, we evaluated not only the accuracy of these participants' mental models, but also the process of constructing understanding, which we did by observing iterations of learning. Evaluating the effects of the external visual materials and domain knowledge allows us to better understand how these function independently, and together, in comprehension.

This study differs from most studies of multimedia learning in three ways. First, we examine the specific mental models, whether correct or incorrect, that people construct from viewing the animations. Following Gentner and Stevens (1983) we define a mental model as a coherent explanation of how a physical system works. Mental models are examined *in addition* to measures of recall and application of knowledge, which are typically evaluated in studies of multimedia learning (e.g., Mayer, 2001). Our measures of knowledge application involved *troubleshooting* (i.e., diagnosing what might be wrong in breakdown situations).

Second, we examine the content of mental models constructed from animations alone, without an accompanying verbal description. We do this for several reasons. First, animations without verbal commentary are often used in technology-based learning materials (e.g., Macaulay, 1998). Second, by not providing verbal descriptions to accompany the animation, we can be sure that performance on written or spoken learning outcomes reflects learners' understanding of the visualization, and does not simply "parrot" information presented in the text. Third, while it has often been shown that people learn more from diagrams that are accompanied by explanatory text or verbal commentaries (e.g., Mayer and Gallini, 1990; Mayer and Anderson, 1991; Hegarty and Just, 1993) one goal of our research is a scientific account of how multimedia presentations are understood, which necessitates isolating the cognitive processes and comprehension outcomes that result from exposure to the individual media that make up multimedia presentations. This scientific account, in turn, can be the basis of a set of principles for the design of multimedia presentations (cf. Narayanan and Hegarty, 1998, 2002), for example by specifying when and how diagrammatic materials should be augmented by text.

The third novel contribution of these studies is to assess the process of learning from animations, and not just the outcomes of this cognitive process. We do this in Experiments 3 and 4 by tracking students' interactions with an interactive animation, tracking their eye fixations as they learn, and allowing students multiple learning iterations with the materials to evaluate how their mental models are constructed and revised with increasing exposure.

2. Experiment 1

In Experiment 1, we compared learning outcomes of people who viewed an interactive animation to those of others who viewed a non-interactive animation. The specific type of interactivity investigated here allowed participants to move forward and backwards in the animation, play it slower or faster, and pause the animation, as in standard video playback computer programs. With such playback features, participants could control the animation to play at a pace that matched the speed of their comprehension processes (cf. Tversky et al, 2002). Participants in the computer-controlled animation condition, on the other hand, did not have the ability to move backwards or pause the animation, and they could only play the animation at a constant rate that may or may not have matched their comprehension speed.

Several researchers have proposed that interactivity can help alleviate difficulties in perceiving and comprehending animations by allowing users to view animations at a pace that is congruent with the speed of their comprehension processes (Tversky et al., 2002; Hegarty, 2004; Schwan and Riempp, 2004). If this is true, participants viewing an interactive animation should have a better understanding of the system than participants who viewed an animation over which they had no control. Thus, we predicted that participants in the interactive condition would perceive and recall more steps in the causal chain of the cistern's cycle and construct more accurate mental models of how the system functions.

In addition to the interactive and non-interactive conditions of the experiment, a control group saw only a static diagram of the toilet tank in resting position during their learning phase. The purpose of this condition was to provide a baseline measure of comprehension of the flushing cistern by participants who are familiarized with a static diagram and the names of the parts, but who receive no explicit information about how the parts move. An obvious additional prediction was that participants in both animation conditions would have superior performance on the comprehension outcomes compared to the control group, because the animations provided information about the motion of the system. A comparison of the animation groups to the control group allowed us to evaluate how much participants learned from the animations.

2.1. Method

2.1.1. Participants

Sixty undergraduate psychology students at University of California, Santa Barbara volunteered to participate in the study for course credit. None had taken college level physics or mechanics courses. The participants were randomly assigned to one of three conditions.

2.1.2. Materials and apparatus

2.1.2.1. Static diagrams. Participants in all three conditions first viewed a labeled static diagram, which showed a color picture of the toilet tank in its resting state and included labels naming the 10 mechanical parts of the system (see Fig. 2). An unlabelled version of the same diagram was shown to control participants during the learning phase of the experiment. The labeled and unlabeled diagrams were viewed as full-screen PowerPoint slides on a 17″ desktop monitor at 1024 × 768 resolution.

2.1.2.2. Animations. Participants in the experimental conditions saw one of two animations, interactive or non-interactive. Both of the animations consisted of a series of 144 bitmap images that illustrated the flushing cistern in motion. The non-interactive animation was displayed in Macromedia Flash MX and played at a rate of 6 frames per second (24 s total play time). The participants pressed a button with the mouse in order to start the animation, but had no control over the speed or the direction in which it played, and could not stop it. The interactive animation was run in Quicktime Player, which allows a person to control the speed and direction in which a video file plays, using a slider bar. In both animations, arrows appeared at various points to indicate a part's direction of movement or to signal an important step.

2.1.2.3. Overview of mechanical system. The flushing cistern has an output and input system, each accomplishing different functions. The output system works to flush water out of the tank and into the bowl. When the handle is pressed down, the disks move together and rise, pushing the water in the siphon bell up and over the curve in the siphon pipe. A siphoning process begins when the pipe fills completely with water. A siphon occurs when liquid in an enclosed system moves, via a pressure differential, from a point of high pressure to a point of low pressure, and in the case of the toilet tank, the siphon process enables water to continuously flow up and down the u-shaped siphon pipe in order to exit the tank. The siphon is broken when the water level in the tank drops far enough that air, which is lighter than water, enters the enclosed system (Fig. 3). As Fig. 3 shows, the disks must not be touching in order to allow air to flow into the siphon bell. The flow of air into the siphon bell and pipe breaks the suction, and the water flow out of the pipe stops immediately.

The inlet system also relies on a physical process to function correctly. As the water leaves the tank and the

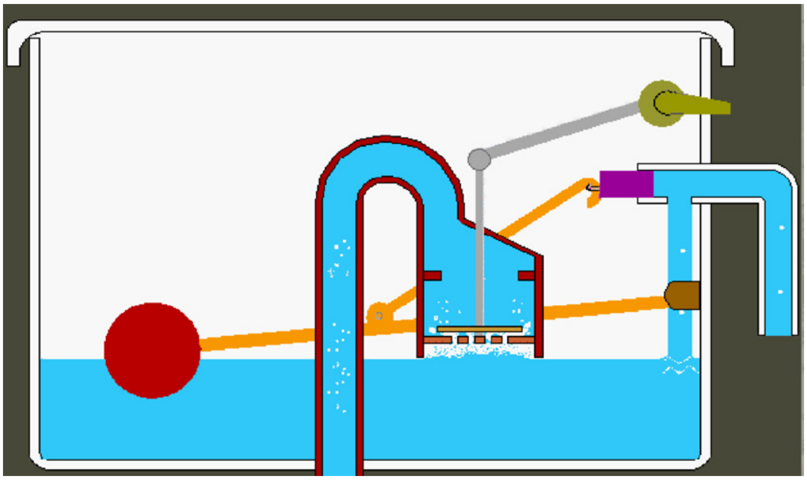

Fig. 3. Frame of the animation showing how air enters and stops the siphon process.

water level lowers, a float, which floats on the water, lowers with the water. The float is attached to an arm, which pulls the inlet valve out. Fresh water enters the tank through this opening. As the tank refills, the float moves up with the water level, and the movement of the float eventually pushes the inlet valve back into place to shut off the flow of fresh water.

2.1.2.4. Comprehension questions. All participants answered a series of written questions. The first question asked them to describe step-by-step what happens in the toilet tank when the handle is pushed. The remaining questions required the participants to troubleshoot the reasons why a breakdown may have occurred. These four questions were:

- Suppose that you push down on the handle of the toilet tank but water does not flush into the toilet bowl. Explain all the possible things that could be wrong.
- Suppose that after flushing the toilet, you notice that water is continuously running into the tank. Explain all the possible things that could be wrong.
- Suppose that after you flush the toilet, water continues to run into the toilet bowl without stopping. Explain all the possible things that could be wrong.
- Suppose that a little while after the toilet has been flushed, water overflows from the top of the toilet tank. Explain all the possible things that could be wrong.

Each question was written on a separate sheet of paper and a small hand-drawn picture of a toilet was shown highlighting whether the question referred to the toilet *tank* or the toilet *bowl*.

2.1.3. Procedure

Participants in all conditions were seated in front of a desktop computer and first studied the labeled diagram for 1 min in order to learn the names of the parts. Then the learning phase began. Participants in the animation conditions were first shown how to manipulate the controls of their animations, and were then shown either the non-interactive animation or the interactive animation. At this phase, participants in the control condition were shown the static unlabelled diagram. All three groups were instructed, "Please study this (animation/diagram) and learn how this mechanical system works. You have as long as you'd like to study this material, so please take as much time as you'd like." Each participant's viewing time was recorded.

After participants viewed the materials, the monitor was turned off and participants were given a booklet of comprehension questions. Seven minutes were allowed for the step-by-step explanation, and 2 min were given for each troubleshooting question. At the end of the experiment, participants answered a questionnaire about their prior mechanical experience as well as providing a self-rating of their interest in the experiment.

2.1.4. Scoring

In order to assess participants' understanding of the steps that occur during a single cycle, the step-by-step responses were coded using a slightly modified version of the coding scheme presented in Hegarty, et al., (2003)[1]. The original coding scheme exhibited 93% agreement between two independent coders on 20 randomly chosen descriptions. One of these two coders scored the present data, while remaining blind to the experimental conditions of the participants.

Participants' step-by-step responses were also coded to qualitatively evaluate the mental models they built. Specifically, responses were coded for correct or incorrect accounts of how the inlet and outlet processes begin and end. Accounts were scored as correct if the following steps

---

[1] The original coding system was modified by deleting four steps because they encoded events that were too finely granular, combining two steps into one, and coding for additional steps that were more diagnostic of participants' mental models. These changes yielded the 17 steps shown in Table 1.

(shown in Table 1) were reported: steps 2, 3, and 5 for the outlet beginning; 6 and 7 for the outlet ending; steps 11, 12, and 13 for the inlet beginning; and 15, 16, and 17 for the inlet ending. Participants who described these processes using terms or phrases different from the ones presented in Table 1 were credited for correct accounts if their descriptions conveyed the appropriate steps. Thus, participants did not have to explicitly use the terms "siphoning" and "buoyancy" as long as they correctly reported the physical events that comprise these phenomena.

Troubleshooting questions were scored by counting the number of plausible reasons for breakdown participants gave. General reasons that did not reflect understanding of the mechanism, such as, "The water in the house was turned off," were not counted.

2.2. Results

2.2.1. General understanding of the system

An ANOVA was conducted on the number of steps mentioned in the step-by-step reports and a main effect of instruction was found ($F(2, 57) = 4.48$, $p = .016$, $\eta^2 = .14$). Tukey post-hoc comparisons indicated that the interactive animation group did not report significantly more steps than the non-interactive animation group. The interactive animation participants reported significantly more steps than the static diagram group ($p = .017$), and the difference between the non-interactive and diagram conditions was marginally significant ($p = .070$). The means and standard errors for each group are shown in Fig. 4.

Table 1
Coding scheme for step-by-step reports

1	Push down on handle	10	Lower disk falls down
2	Disks move up	11	Water level lowers
3	Water goes through siphon bell/pipe	12	Float lowers
4	Water enters toilet	13	Inlet valve opens
5	Siphon process begins	14	Water flows into tank
6	Water level falls below siphon bell	15	Water level rises
7	Siphon process is broken	16	Float rises
8	Handle is released	17	Inlet valve closes
9	Upper disk falls down		

Steps 1–10 refer to actions of the outlet system and steps 11–17 refer to actions of the inlet system.

Steps reported by at least 70% of participants from each group are shown in bold numbers in Table 2. While there are slight differences in the steps that were reported by the different groups, there is a clear pattern of responses across all three groups. Participants generally mentioned the first three steps of the outlet process and then moved to focus on the inlet process.

2.2.2. Accuracy of mental models

Step-by-step responses were also evaluated specifically for mention of the beginning and ending of both the inlet and outlet processes. The results are shown in Table 3. Most participants, regardless of condition, failed to mention the beginning and ending of the outlet process. Of the few who did explain the ending of the outlet process, most did not explain it correctly. The majority of these participants reported that the disks stopped the water from flowing into the siphon pipe, rather than reporting that the air enters and the siphon process ends. The participants were much more consistent in reporting the correct beginning and ending of the inlet process. There was a significant effect of condition in the reporting of the ending of the inlet process ($\chi^2(2, N = 60) = 8.4$, $p = .015$). However, this reflects a difference between the diagram and animation conditions, and does not suggest any benefit of interactivity.

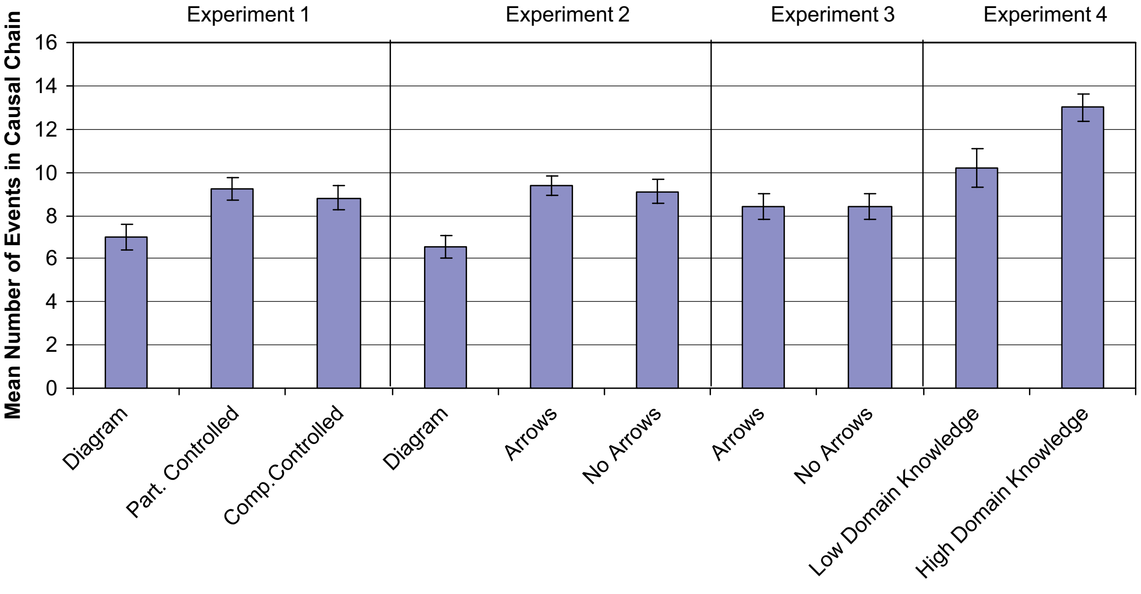

Fig. 4. Mean number of steps in the causal chain mentioned by each group in Experiments 1–4.

Table 2
Proportions of participants in each experimental condition who reported each step of the causal chain

Step	Exp 1			Exp 2			Exp 3		Exp 4	
	Diagram	Inter-active	Non inter-active	Diagram	Arrows	No arrows	Arrows	No arrows	LDK	HDK
1	***1***	***1***	***1***	***1***	***1***	***1***	***1***	***1***	***1***	***1***
2	***.85***	***.9***	***.7***	***.85***	***.7***	***.7***	***.8***	***.7***	***1***	***1***
3	***.75***	***.9***	***.95***	***.8***	***.8***	***.95***	***.8***	.6	***.89***	***1***
4	.35	.2	.2	.25	.2	.1	.1	.2	.22	.6
5	.1	.15				.15	.1			.5
6			.1						.11	.6
7			.15						.11	.2
8	.15				.1		.1		.33	.6
9	.2	.45	.3	.2	.55	.5	.6	.6	***.89***	.5
10		.25	.25		.5	.4	.1	.4	.33	***.7***
11	.55	***.85***	.55	.35	***.7***	***.75***	.5	***.8***	.67	***.9***
12	***.7***	***.85***	***.75***	***.7***	***.95***	***.75***	***.9***	***.7***	***.78***	***.9***
13	***.7***	.65	***.7***	.55	***.7***	***.7***	***.7***	.5	***.78***	***1***
14	***.75***	***.8***	***.85***	***.8***	***1***	***.9***	***.9***	***.9***	***.89***	***1***
15	.25	***.8***	.55	.5	***.75***	***.8***	.4	.6	.56	***.8***
16	.35	***.8***	***.85***	.25	***.75***	.65	***.8***	***.7***	***.78***	***.7***
17	.25	.6	***.85***	.3	***.7***	.45	.6	.5	***.89***	***1***

Proportions of .70 or higher are printed in bold and italics. Blank cells represent steps with proportions of .05 or less. Experiment 4 data reflect the first learning iteration only.

Table 3
Proportions of participants in Experiment 1 who mentioned correct or incorrect sub-system processes in the causal chain responses

	Static diagram	Interactive	Non-interactive
Outlet beginning			
Correct	.10 (2)	.15 (3)	.10 (2)
Incorrect	.00 (0)	.00 (0)	.00 (0)
Inlet beginning			
Correct	.55 (11)	.60 (12)	.70 (14)
Incorrect	.00 (0)	.00 (0)	.00 (0)
Outlet end			
Correct	.00 (0)	.00 (0)	.10 (2)
Incorrect	.05 (1)	.25 (5)	.20 (4)
Inlet end			
Correct	.25 (5)	.55 (11)	.70 (14)
Incorrect	.00 (0)	.00 (0)	.00 (0)

Numbers in parentheses represent actual number of participants.

2.2.3. Troubleshooting

For each troubleshooting question participants were asked to generate possible reasons why the breakdown scenario presented could have occurred. The number of correct reasons generated across all four questions was calculated and an ANOVA revealed that there were no significant differences between the learning groups ($F(2,\ 57) = 2.26$, $p = .11$, $\eta^2 = .07$). Fig. 5 illustrates the means and standard errors of each group.

2.2.4. Learning time

The amount of time participants spent engaging with the learning materials differed across conditions. Both the participant-controlled animation group ($M = 195$ s, SD = 104 s) and the computer-controlled animation group ($M = 140$ s, SD = 57 s) took a significantly longer amount of time to view their materials than the static diagram group ($M = 74$ s, SD = 49 s, $F(2,\ 57) = 13.36$, $p = .00$, $\eta^2 = .32$). There was also a trend for the participant-controlled animation group to spend more time on the materials than the computer-controlled animation group ($p = .06$).

2.3. Discussion

The results from Experiment 1 indicate that participants who viewed an interactive animation did not obtain an enhanced understanding of the flushing cistern when compared to participants who received the same information presented in an animation without speed and direction controls. Moreover, the data provide no support for the prediction that participants viewing the interactive animation construct more accurate mental models of the inlet and outlet systems. Because participants in this condition could stop the animation at particular points and inspect the fine details of the animation, we predicted that they would gain a better understanding of how the flushing cistern works. However, there was no indication that participants in the interactive group constructed more accurate mental models than participants in the computer-controlled animation group.

One possible reason for the lack of difference between the two animation groups might be that neither animation group learned anything from the animations. However, both groups reported more steps in the causal chain than the unlabeled static diagram group, who viewed a single

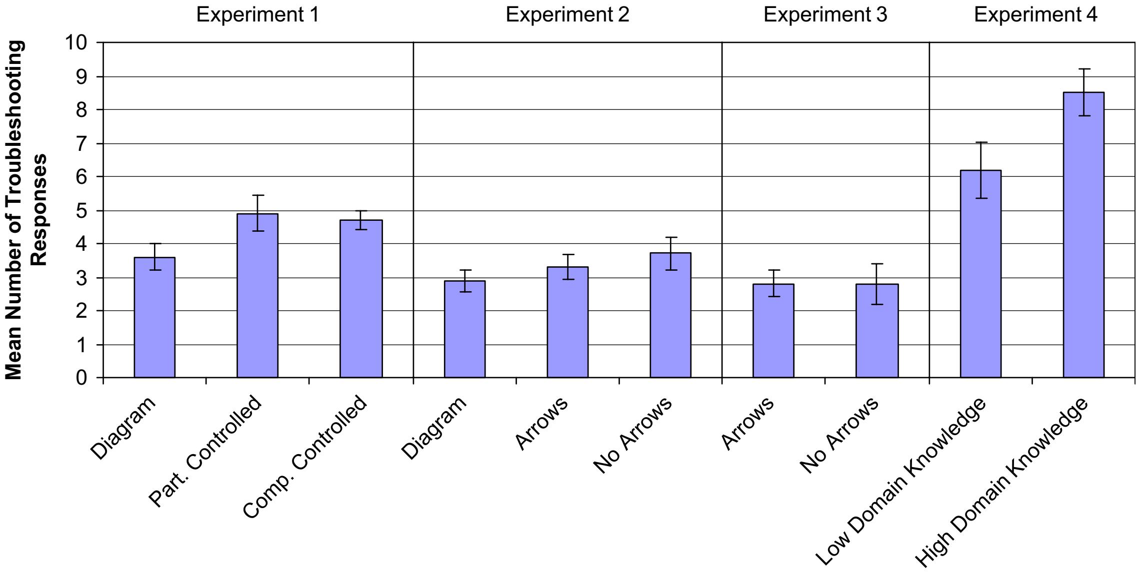

Fig. 5. Mean number of troubleshooting responses generated by each group in Experiments 1–4.

diagram showing the components of the system, but not the motion of the parts. This suggests that participants gained some knowledge from the animations, and the lack of statistically significant differences between two animation groups was not simply due to a floor effect.

In considering the qualitative features of participants' mental models, it seems that the animation groups developed a general understanding of the inlet sub-system of the flushing cistern. However, only two participants across all of the conditions reported the correct mental model of the outlet system, and 10 participants reported the same erroneous model of this system, that the disks stop the flow of water. Note that this explanation is inconsistent with the visual display, because the display showed that the disks were separated when the outflow of water stops (see Fig. 3). One explanation for this difference in understanding of the inlet and outlet systems may be that these students are applying their knowledge of an American toilet tank while viewing learning materials or when answering the comprehension questions. The American model and the model depicted in the animations are similar with respect to the inlet system, but in an American toilet tank, the closing of a rubber flap stops the outflow of water. An alternative explanation for why participants developed an erroneous model of the outlet system is that the physical process of siphoning is more complicated and may not be observed as often by physics novices as the physical phenomenon of buoyancy.

A limitation of Experiment 1 is that participants' mental models were evaluated based on their step-by-step descriptions, and when people failed to mention a step, it was not possible to tell whether their mental model was correct, incorrect, or incomplete. Thus, in the remaining experiments, we asked people focused questions (function questions) in which we explicitly asked them what stops the inlet and outlet systems. Their answers to these questions may draw on a complete mental model of the system that was constructed while viewing the animation, or an incomplete or implicit understanding that was only filled in or made explicit when answering the question. Nevertheless, they are a reflection of the participants' understanding of the system after their various learning experiences.

The significant differences in learning time may have affected the results obtained in this experiment. That is, the superior comprehension of the animation groups may have been merely due to the fact that they spent more time on the materials. This seems unlikely considering that the two animation groups differed in learning time but not on any comprehension measures. Nevertheless, we avoided this confound in Experiment 2 by controlling for study time.

3. Experiment 2

In Experiment 2, we evaluated the effect of signaling on constructing understanding from an animation. Previous research (e.g., Grant and Spivey, 2003; Mayer and Moreno, 2003) suggests that animations might be made more effective by including signaling devices that draw the learner's attention to the relevant parts of the external display. This may be particularly important for complex animations in which several components are moving at once, and learners need to focus on a subset of the components, for example to understand a causal sequence of events (Narayanan and Hegarty, 1998). Both animations in Experiment 1 contained signaling devices, and these devices may have contributed to participants' understanding of the flushing cistern. In Experiment 2 we compared the interactive animation used in Experiment 1, which contained arrows as signaling devices, to the same animation without arrows. As in Experiment 1, a third group of participants served as a control group, and it was predicted that they would have poorer performance because they were not shown the movement of the system.

If signaling devices are effective in directing attention to relevant steps, participants who view the animation with arrows should build a more complete understanding of the system. Thus, we predicted that participants who viewed the animation with arrows would report more steps in the causal chain and better understand the correct ending processes of the inlet and outlet sub-systems than participants who did not receive signaling devices. Furthermore, if arrows successfully draw attention to the steps they signal, participants viewing the animation with arrows should mention these steps more often than participants who did not receive signaling devices. Specifically, participants who viewed the animation with arrows should report the endings of the inlet and outlet sub-cycles, because the ending of these two processes are signaled by arrows.

3.1. Method

3.1.1. Participants

Sixty undergraduate psychology students at the University of California, Santa Barbara volunteered to participate in the study for course credit. None had taken college-level physics or mechanics courses, and none of the participants had taken part in Experiment 1. All participants were randomly assigned to one of the three conditions.

3.1.2. Materials and apparatus

The labeled and unlabeled static diagrams were identical to those used in Experiment 1. The interactive animation with arrows from Experiment 1 was also used. A second interactive animation, without arrows, was shown to one-third of the participants. In the animation with arrows, arrows appeared during eight steps of the flushing cistern's cycle: (1) when the handle is pressed down, (2) when the disks rise up, (3) when the water flows through the siphon pipe, (4) when the float moves down, (5) when the inlet valve moves out, (6) when air enters the siphon bell, (7) when the float moves up, and (8) when the inlet valve moves in. Each arrow flashed on and off for the duration of its step. Two types of arrows were used, arrows showing a direction of movement (e.g., how water flows up the siphon pipe and out of the tank) and arrows pointing to the occurrence of a step (e.g., an arrow appearing at the siphon bell at the point that air enters).

The comprehension measures were the same as those used in Experiment 1. In addition, four function questions were added. Two of these questions, "What stops water from exiting the tank?" and "What stops water from entering the tank?" were included specifically to assess participants' mental models of the stopping mechanisms in both systems[2].

3.1.3. Procedure

The procedure of Experiment 2 was identical to Experiment 1 except that data were collected in a larger room, accommodating up to three participants per session. Additionally, participants in all conditions were given 3 min to study the learning materials. They could not move on to the comprehension questions until the 3 min had passed.

3.2. Results

3.2.1. General understanding of the system

The effect of learning condition on the number of steps reported in the causal chain descriptions was significant ($F(2, 57) = 9.02$, $p = .00$, $\eta^2 = .24$). Tukey post-hoc comparisons indicated that both the group who viewed the animation with arrows ($p = .01$) and the no-arrows group ($p = .01$) reported significantly more steps in their step-by-step responses than the static diagram group. This indicates that participants in both animation groups learned more than participants studying a single static diagram. However, as Fig. 4 illustrates, the number of steps reported by participants in the animation groups did not differ statistically, suggesting that the arrows did not cause participants to incorporate more steps into their descriptions of the system. Furthermore, Table 2 shows little difference between the two animation conditions in the specific steps mentioned by the majority of participants in each group. However, the majority of the static diagram group reported fewer steps than the animation groups.

In a more focused analysis, we examined comprehension of the eight steps in the causal chain that were specifically signaled by the arrows. Participants' responses were coded specifically for these eight steps, and the number of steps mentioned was compared across the two animation conditions. Planned comparisons revealed that the difference between the two animation groups was not significant ($t[38] = .77$, $p = .45$, $\eta^2 = .02$). The animation with arrows group reported an average of 5.60 of these steps (SD = 1.39), while the animation without arrows group reported a mean of 5.25 (SD = 1.48).

3.2.2. Accuracy of mental models

An evaluation of model types was conducted by evaluating answers to function questions regarding the ending of the inlet and outlet processes. Table 4 reports the frequency of correct and incorrect responses for these steps for each of the experimental groups, indicating that regardless of condition, participants were able to answer the question about the inlet process correctly, while they tended to answer the outlet question incorrectly. Wilcoxon

---

[2] Responses to the two other function questions are not discussed here, as they did not differentiate between the groups in any of our experiments.

Table 4
Proportions of correct and incorrect responses to the outlet and inlet function questions

Condition	Outlet correct	Outlet incorrect	Inlet correct	Inlet incorrect
Experiment 2				
Diagram	.00 (0)	.80 (16)	.65 (13)	.25 (5)
Arrows	.00 (0)	1 (20)	.95 (19)	.00 (0)
No arrows	.05 (1)	.90 (18)	.70 (14)	.20 (4)
Experiment 3				
Arrows	.00 (0)	1 (10)	.90 (9)	.10 (1)
No arrows	.00 (0)	.90 (9)	.60 (6)	.30 (3)
Experiment 4				
LDK	.00 (0)	1 (9)	.89 (8)	.11 (1)
HDK	.40 (4)	.30[a] (3)	1.0 (10)	.00 (0)

Actual number of participants is shown in parentheses. Experiment 4 data reflects the first learning iteration only.

[a] Three of the HDK participants gave ambiguous answers to the outlet question, so it was not possible to determine whether they had the correct or the erroneous mental model.

signed rank tests performed separately on the data for each group revealed that all three groups understood the inlet system more than the outlet system. Thus for the static diagram group ($Z = 3.61$, $p = .00$), no arrows group ($Z = 3.32$, $p = .01$), and arrows group ($Z = 4.359$, $p = .00$), the inlet question was correctly answered significantly more often than the outlet question. In testing for differences between conditions, a χ^2 test yielded no significant differences between the three groups in their responses to the outlet ($\chi^2(4, N = 60) = 7.64$, $p = .11$) or inlet ($\chi^2(4, N = 60) = 6.41$, $p = .17$) questions.

A qualitative analysis of the incorrect answers revealed that the majority of participants, regardless of condition, reported that the disks stopped the outlet process. Of the incorrect outlet responses, 72% mentioned the disks as causing the stopping of water flow out of the tank. Similarly, 67% of the incorrect inlet process responses named the disks as responsible for stopping the inlet process. These qualitative analyses suggest that the function of the disks is not well understood by participants, regardless of the type of visual materials viewed.

3.2.3. *Troubleshooting*

No significant group differences were found in the number of troubleshooting responses generated. As Fig. 5 shows, although the static diagram learners generated fewer reasons for breakdown ($M = 2.85$, SD = 1.496) than the animation with arrows ($M = 3.25$, SD = 1.618) and animation without arrows groups ($M = 3.70$, SD = 2.179), the difference was not significant ($F(2, 57) = 1.130$, $p = .330$, $\eta^2 = .038$).

3.3. *Discussion*

The results of Experiment 2 indicate that signaling devices did not contribute to a better understanding of the flushing cistern. Specific steps that were signaled were not mentioned more often by participants who saw the arrows compared to people who did not see the arrows, and answers to the function questions were similar for participants who received signaling and those who did not. However, both animation groups reported significantly more steps in the causal chain than the static diagram group, suggesting that participants learned from the animation, and the finding of no difference between the animation conditions was not simply a floor effect.

Surprisingly, the participants were extremely consistent in the erroneous mental models they reported, and this was true in all three conditions. When questioned directly, the vast majority of all participants had the misconception that the disks serve as the stopping agents for the outlet process, and many participants erroneously named the disks as the cause of the ending of the inlet process as well.

The combined findings from Experiments 1 and 2 suggest that animations designed to alleviate either a mismatch between the pace of comprehension and the pace of an animation, or a mismatch between visual salience and relevance do not lead participants with low physics knowledge to superior understanding of the flushing cistern. One possibility for these null findings is that the participants may not have utilized the interactive or signaling features of the animations (cf. Pane et al., 1996; Lowe, 2004). Alternatively, participants may have utilized the devices, but may have done so in an ineffective manner. Because Experiments 1 and 2 were not designed to collect data on levels of interactivity or utilization of signaling devices, we conducted a separate experiment to test whether these manipulations were utilized.

4. Experiment 3

4.1. *Method*

4.1.1. *Participants*

Twenty undergraduates taking an introductory psychology class participated in the experiment for course credit. None had taken college level physics or mechanics classes or had participated in Experiment 1 or 2. All participants had normal or corrected-to-normal vision.

4.1.2. *Materials and apparatus*

The animations were the same as those used in Experiment 2, except that they were controlled by pressing the left and right arrow keys on the keyboard, rather than by using the mouse and slider bar. Use of this interface was necessary so that we could simultaneously monitor participants' eye fixations and frame display times.

The animations were presented on a 15″ monitor with a screen resolution of 800 × 600. Participants were seated approximately 20″ away, so that the visual angle of the screen was approximately 38° × 27°. Their eye fixations were measured using a SensoMotoric Instruments Eyelink head-mounted eye tracker that sampled eye fixations in

x and y coordinates at a rate of 250 Hz with an angular error of less than .5°. Data were output to a PC computer for further analysis. The comprehension booklet was the same one used in Experiment 2.

4.1.3. Procedure

Participants were first fitted with the head-mounted eyetracker, the system was calibrated, and an unrelated experiment was run. Upon completion of the other experiment, the participant viewed the labeled static diagram of the toilet tank for 1 min, to become familiar with the part names. The eyetracker was then re-calibrated and the participants viewed either the interactive animation with arrows or the interactive animation without arrows. Participants were given an unlimited amount of time to view the animation, and when they indicated they were finished, the head mounted cameras were removed and participants answered the comprehension questions with the same time limits as in Experiment 2.

4.2. Results

4.2.1. General understanding of the system

The mean number of steps in the causal chain mentioned by the two groups was similar to the results for the corresponding groups in Experiment 2. There was no significant difference between conditions in the number of steps reported (See Fig. 4), and examination of the steps mentioned by 70% or more of participants showed a similar pattern to Experiments 1 and 2 (see Table 2).

4.2.2. Accuracy of mental models

The number of correct and incorrect responses to the inlet and outlet function questions also exhibited similar patterns to those of the animation participants in Experiment 2. Table 3 shows that no participants correctly answered the outlet question, whereas most participants answered the inlet question correctly. Wilcoxon signed rank tests revealed a significant trend in both groups such that the inlet question was answered correctly more often than the outlet question (Arrows: $Z = 3.00$, $p = .01$; No Arrows; $Z = 2.45$, $p = .01$). No significant group differences were found for the outlet question ($\chi^2(1, N = 40) = 1.05$, $p = .31$) or the inlet question ($\chi^2(2, N = 40) = 2.60$, $p = .27$).

Qualitative analyses suggest that these participants, like those in Experiments 1 and 2, attributed the disks falling as the reason for the ending of the outlet process. Of the incorrect outlet answers, 79% reported the disks, indicating that the majority of the participants in this study constructed the same incorrect mental models of the outlet system.

4.2.3. Troubleshooting

As in Experiment 2, no significant differences were found in the number of correct troubleshooting responses generated by these groups. As Fig. 5 shows, both groups generated a mean of 2.80 responses (SD = 1.23 for the animation with arrows group, SD = 1.93 for the animation without arrows group).

4.2.4. Interactivity

Interactivity data for each participant were plotted in a scatterplot to visually examine the movement through frames. In these plots, the x-axis represents time, and the y-axis indicates the frame viewed at each point in time. A peak indicates that the participant switched from moving forward in the animation to moving backward. A common strategy (exhibited by 11 of the 20 participants) was to view the entire animation and then return to the beginning. However, there was individual variability in when participants chose to move backwards in the animation and how many times they moved backwards to review some of the frames. To illustrate, the interactivity data from four representative participants are plotted in Fig. 6. Of these four participants, two (Participants 16 and 2) show fairly non-interactive profiles. They merely viewed the animation from start to finish a couple of times. The other two (Participants 14 and 11) were highly interactive. As the plots show, they reviewed specific parts of the animation as well as the whole animation several times.

We computed two measures of interactivity. The first was the number of times a participant moved backwards in the animation. Combining the data from both groups, the mean number of times participants moved the animation backwards was 6.3 (range 1–18). The second was the standard deviation of the amount of time spent viewing each frame. Theoretically, the standard deviation of the viewing time of each frame in a non-interactive animation is zero, because each frame is visible for the same amount of time. No participant in this experiment had a standard deviation of close to zero (the standard deviation scores ranged from 506 to 3813 ms) indicating that all participants made use, to some extent, of the variable speed at which the animation could be viewed.

Neither interactivity measure was found to correlate significantly with any of the comprehension measures (all correlations were less than .2), and there were no significant group differences in levels of interactivity, either in backward movements (Arrows: $M = 6.70$, SD = 5.56 vs. No Arrows: $M = 5.90$, SD = 5.15 ($t[18] = .33$, $p = .74$, $\eta^2 = .01$)) or in standard deviations of frame time (Arrows: $M = 1433$ ms, SD = 540 ms vs. No Arrows: $M = 1800$ ms, SD = 1144 ms, $t[18] = -.92$, $p = .37$, $\eta^2 = .05$).

The comprehension data from this and the previous experiments suggests that students have particular difficulty understanding the end of the outlet (siphoning) process. One possibility is that participants quickly skip over this process, so they do not adequately view this part of the animation. To evaluate this possibility, we computed the average time per frame spent on frames 63–71, which showed the end of the outlet process and the average time across all other frames. In other words, this analysis compared time average time spent looking at the steps that

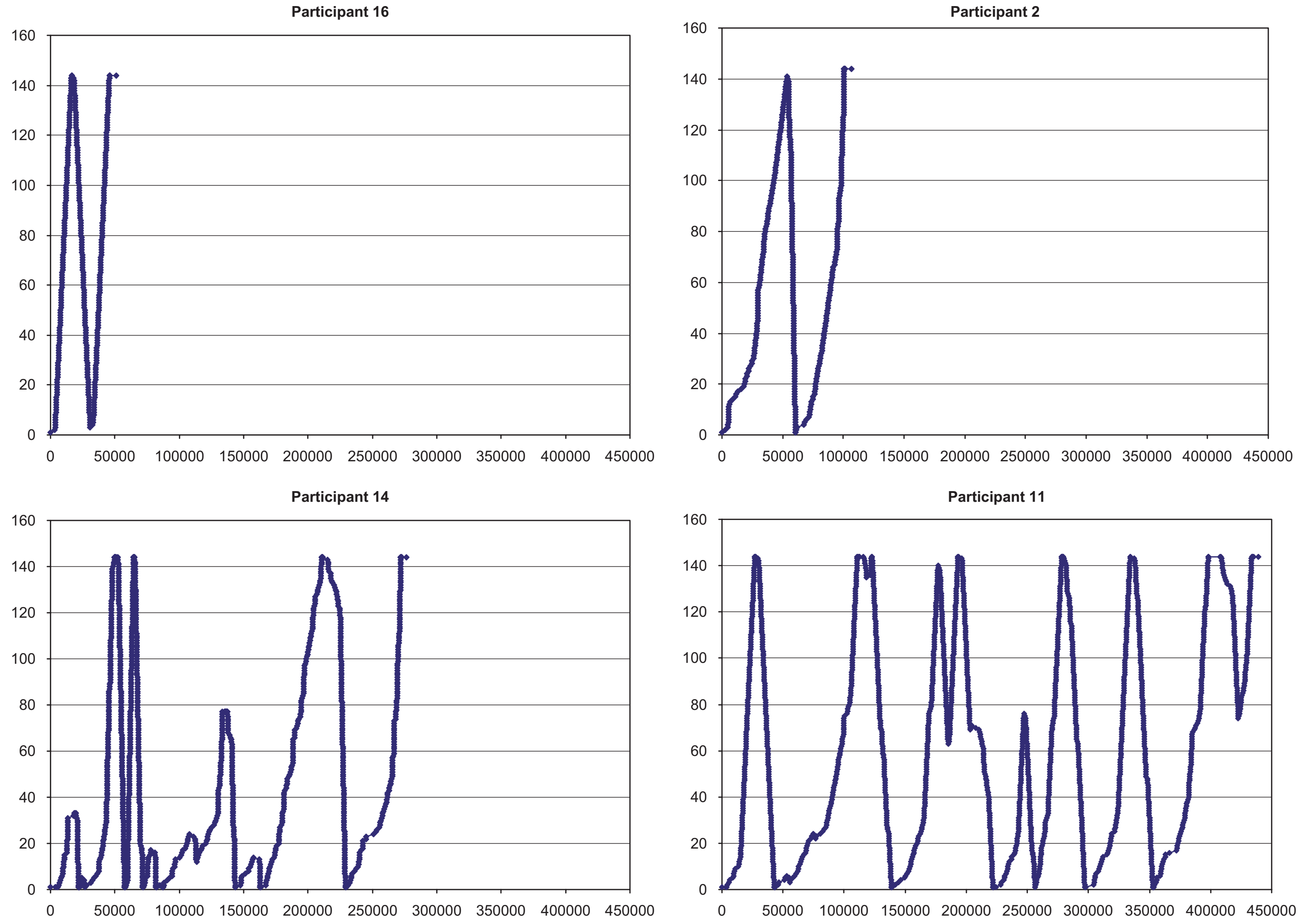

Fig. 6. Interactivity plots of four participants.

ended the outlet process to all other steps (every frame of the animation conveyed the occurrence of at least one step). Participants spent significantly more time per frame viewing frames showing the end of the outlet process ($M = 1559$ ms, SD = 1087) than they spent on the frames depicting other steps ($M = 1083$ ms, SD = 605, $t[19] = 2.65$, $p = .02$, $\eta^2 = .27$). Average viewing time of frames depicting the outlet ending process were not different for the animation with signaling ($M = 1364$ ms, SD = 848 ms) and non-signaling animation groups ($M = 1755$ ms, SD = 1300 ms; $F(1, 18) = .64$, n.s., $\eta^2 = .03$) and average viewing time during this segment of the animation was not significantly correlated with any of the comprehension measures (all correlations were less than .3).

4.2.5. *Signaling*

The eye-fixation data were analyzed to determine if participants viewing the signaling animation looked more often at the parts being signaled than the participants who did not receive arrows. Two participants from the signaling condition were eliminated from this analysis because the eye tracking system lost calibration during the experiment, so this analysis was based on data from the remaining 18 participants.

To analyze the eye movement data, we first divided the animation into "episodes" or series of consecutive frames that showed a step in the causal chain in which one or more signaling arrows appeared. We then defined "regions of interest" in the visual display for each of these episodes, including: (1) regions of space in which arrows appeared, (2) regions of space showing a component to which an arrow pointed and (3) larger regions that included both an arrow and the part to which it pointed. There were six different episodes, (i.e., series of frames) in which arrows appeared. Across these 6 episodes, a total of 27 regions of interest were defined (in some cases there was more than one arrow). For example, Fig. 7 shows the three regions of interest for the series of frames (frames 63–71) showing the end of the siphoning process; one region surrounding the arrow that appears, another surrounding the siphon bell and disks to which it points, and the third incorporating both of the regions.

Eye fixation samples that occurred during Frames 63–71 are shown in Fig. 7 as small dots (with a sampling rate of 250 Hz, one sample represents 4 ms.). To analyze the effects of signaling quantitatively, the number of eye movement data samples that occurred in these regions within the specified series of frames was counted and this number was expressed as a percentage of total number of eye movement samples that occurred during the frame series (episode). This yielded a measure of percent of time on the "signaled" regions of interest that could be compared across the two groups.

Independent samples t-tests revealed that participants receiving the animation with arrows spent a significantly greater proportion of time in the arrow regions ($M = 5.8\%$, $SD = 2.68$) and the space that incorporated both the parts and arrows ($M = 22.31\%$, $SD = 7.82$) than the participants who saw the animation without arrows (respectively, $M = 3.42\%$, $SD = 2.00$, $t[16] = 2.24$, $p < .05$ and $M = 14.16$, $SD = 6.46$, $t[16] = 2.42$, $p = .03$). There was also a non-significant trend for participants in the arrow condition to spend more time on the parts indicated by the arrows (Arrows: $M = 15.63\%$, $SD = 6.50$ vs. No Arrows: $M = 10.86$, $SD = 5.74$, $t[16] = 1.65$, $p = .12$). Fig. 7 illustrates these results by showing that a participant who received the animation with arrows had more data samples in the regions of interest than a participant who received the animation without arrows.

4.3. Discussion

Some previous studies in the literature indicate that people do not always utilize interactive controls when they are provided with interactive animations (Pane et al., 1996; Lowe, 2004). Contrary to these findings, the data from the present experiment indicate that participants did indeed utilize the interactive controls of their animations. Furthermore,

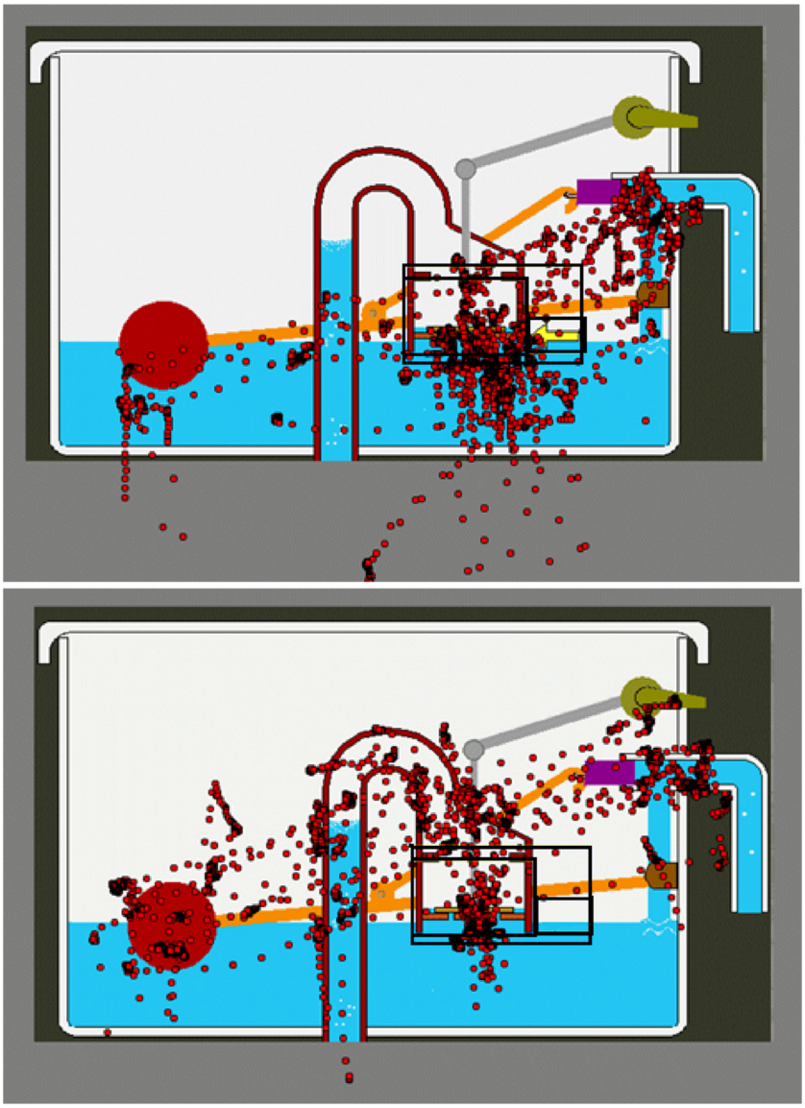

Fig. 7. Example of eye-fixation samples for two participants. The top diagram shows the data of a participant who received the animation with arrows. The bottom diagram shows the data of a participant who saw the animation without arrows appearing. The boxes around the siphon bell and arrow show the regions of interest.

participants who saw the animation with arrows looked more often at the regions of the display highlighted by the arrows than the participants who did not receive this signaling. However, these participants did not perform better on comprehension measures. These results rule out the possibility that the null effects found in Experiments 1 and 2 occurred simply because participants were not utilizing the enhanced features of the animations. What they suggest is that signaling devices can get learners to look in the right place, but this is not sufficient to construct an accurate understanding of the information being depicted. This result makes it clear why it is important to make a distinction between the perceptual processes of extracting the visual features of a display and the more conceptual processes of encoding that display and constructing a mental model of the referent (see Fig. 1). That is, perception and attention to the relevant location in a display do not necessarily imply that the information at that location will be fully comprehended.

The cumulative findings of Experiments 1–3 indicate that interactivity and signaling modifications of animations do not enable these learners to construct an adequate understanding of the system. Although these have been proposed as features of good animations (Tversky et al., 2002; Hegarty, 2004), and some support for them has been shown in previous studies (e.g., Faraday and Sutcliffe, 1997; ChanLin, 1998; Mayer and Moreno, 2003; Schwan and Riempp, 2004), adding these features to a realistic animation does not guarantee more effective learning. Furthermore, the regularity with which the participants in these experiments attribute the stopping of the outflow of water to the disks indicates that, regardless of the visual materials provided, participants often construct erroneous models of the outlet sub-system. The interactivity data from this experiment show that participants spend ample time looking at the ending of the outlet process, yet simply spending time viewing this part of the animation does not lead them to build a correct model of the system. As noted above, the erroneous disk-stopping model that many participants build is inconsistent with the visual display, because the display showed that the disks were separated when the outflow of water stops and therefore the disks could not have stopped the flow of water (see Fig. 3). The similarity of students' incorrect models suggests that some factor besides perception of the external display (i.e., bottom-up processing) was consistently affecting their mental models. Possible shared influences on these models include the participants' prior knowledge (e.g., of American toilet tanks) and their naïve physics intuitions. In Experiment 4, participants with different levels of prior domain knowledge were given identical visual materials in order to explore how prior knowledge and intuitions may contribute to constructing understanding from animations.

5. Experiment 4

Whereas Experiments 1–3 evaluated the role of bottom-up influences on comprehension by manipulating aspects of the display (i.e., the perceptual input) while controlling participants' prior knowledge, Experiment 4 evaluates how domain knowledge affects the mental models constructed from an animation. Assuming that inferring the causal and functional events from visual, non-verbal materials involves top-down processes such as integrating prior knowledge with the information provided in the external display, we expected students' levels of domain-related knowledge to affect their mental models.

In Experiment 4, we compared comprehension of the visual materials by high domain knowledge (HDK) individuals (engineering graduate students) and LDK individuals (graduate students in humanities and social sciences). The HDK participants had at least 5 years of education in physics or engineering in which they had learned about relevant physics principles such as pressure differentials and siphoning. The LDK individuals were assumed to not have this knowledge. Neither group had specific prior knowledge about how the flushing cistern depicted in our animation works.

Experiment 4 follows the same basic design as Experiments 1–3, however, we gave participants access to all of the instructional materials used in the previous experiments including all the animations and static diagrams used in Experiments 1–3, in order to assess the maximum level of understanding that can be gained from access to these materials. We also included a diagram illustrating four key phases of the flush cycle, which was found to be informationally equivalent to a non-interactive animation in a previous study (Hegarty et al., 2003). In this experiment, we asked participants to view the visual materials a second time after answering the written comprehension questions. Thus, participants initially viewed the materials, created a mental model of the system, described this mental model in writing and then were given a chance to compare their internal models, which they had made explicit in writing, with the external visualizations of the system. During the second viewing, they were asked to report the steps that occur in the cistern's cycle orally, while moving through the animation. This additional task enabled us to assess how mental models of a mechanical system are initially built and how they are revised, as learning continues.

Assuming that constructing understanding from animations is an iterative process (Miyake, 1986), how people move from a state of non-understanding to a state of understanding may depend on their level of domain-related knowledge. As previous studies have shown (Chi et al., 1994; Chi, 2000), it is when conflicts between internal models and external information occur that people are more likely to revise their internal mental models. We proposed that conflicts between internal and external models would be detected more often by people with HDK than by those with LDK, and therefore, HDK individuals would be more likely to revise their mental models. Furthermore, we were interested in evaluating qualitative differences between the specific mental models of the flushing cistern constructed by learners with high and low physics knowledge. Because HDK participants have more relevant prior knowledge that can be applied to construct a mental model from the external display (see Fig. 1), we predicted that these participants would be less likely to report that the disks stop the outflow of water, and instead would report the correct siphon process.

5.1. Method

5.1.1. Participants

Nineteen adults (10 HDK, 9 LDK) volunteered for the study as paid participants. The HDK participants were graduate students in Mechanical Engineering or Materials Science, with the exception of one participant, who was the staff lab manager of the undergraduate Mechanical Engineering Laboratory. All HDK participants held Bachelors degrees in engineering or physics and had been studying physics and engineering for at least 5 years (Mean = 6.4).

The LDK participants were UCSB graduate students from Social Sciences, Art, Humanities, or Biology, and all considered themselves physics novices. Four had taken introductory physics in high school and one had taken introductory physics in college. None had taken engineering courses.

5.1.2. Materials and apparatus

The materials were the same as those used in Experiments 1 and 2 and included the static labeled diagram (Fig. 2), the static unlabeled diagram, the non-interactive animation, the interactive animation without arrows, and the interactive animation with arrows. An additional static diagram was used, which showed the toilet tank at four different phases of the flush cycle. This diagram, like the other static diagrams, was displayed as a PowerPoint slide. All materials were displayed on a 17″ desktop monitor at 1024 × 768 resolution.

5.1.3. Procedure

Participants sat in front of the computer and were given access to all of the visual learning aids. The researcher briefly explained each learning aid, without mentioning the presence or absence of arrows in the animations, and demonstrated how to manipulate the controls of the Quicktime Player. Participants were told to take as much time as they needed to view the materials with the goal of understanding how the toilet tank worked. During this learning phase, the researcher watched participants interact with the materials and ensured that they spent time viewing each of the diagrams and animations. Participants spent approximately 3–5 min studying the materials.

After viewing the materials, the monitor was turned off and participants were given a booklet of comprehension questions. These questions were the same as those used in Experiments 2 and 3. Participants provided written answers at their own pace.

After answering the written questions, participants were asked to view the visual materials again. They were given access to all the visual materials and were asked to think about where events began and ended. An "event" was not predefined by the researcher, and participants were allowed as much time as they needed to accomplish this task. When they were ready, participants reported what they saw as "events" to the experimenter, using the interactive animation with arrows. This portion of the session was video and audio taped for later analysis.

5.2. Results

5.2.1. General understanding of the system

As Fig. 4 shows, the HDK participants reported significantly more steps in their written causal chain responses than the LDK participants after the first learning phase ($F(1, 17) = 6.64$, $p = .02$, $\eta^2 = .28$). Table 2 also shows some differences between the groups in the specific steps of the causal chain that were reported. For instance, several of the outlet steps (such as the beginning of the siphon process or the water level falling below the siphon bell) were reported by at least half of the HDK participants, whereas the LDK participants, like the participants in Experiments 1–3, rarely mentioned these steps.

5.2.2. Accuracy of mental models

As shown in Table 4, HDK participants were more likely than LDK participants to correctly answer the function question regarding how the outlet process ends after the first learning phase ($\chi^2(2, N = 19) = 7.89$, $p = .02$). The HDK group exhibited more variability in the type of explanations they provided, whereas the LDK group uniformly reported the predominant erroneous mental model. No significant differences were found between the groups in their response to the inlet function question ($\chi^2(1, N = 19) = 1.17$, $p = .28$). Both LDK and HDK participants were more likely to correctly answer how the inlet process ends than how the outlet process ends (LDK: $Z = 2.83$, $p = .01$; HDK: $Z = 2.00$, $p < .05$).

5.2.3. Mental model revisions

Answers to the function questions were compared to the oral description of the process given by each participant while viewing the materials for the second time, in order to assess how the mental models of the outlet system changed with the learning process. This comparison focused on participants' explanations of how the outlet system ends, and is presented in Table 5. As the table indicates, most HDK participants, by the end of the experiment, reported an accurate model of the output system, which sometimes involved revising or elaborating their initial mental models. The LDK participants, on the other hand, did not revise their models. The incorrect mental model was given by a majority of these participants in both their initial and final descriptions.

Table 5
Summary of initial (based on answer to the function questions) and final (oral response) understanding of outlet ending process for participants in Experiment 4 with high domain knowledge and low domain knowledge

	High domain knowledge group			Low domain knowledge group	
	Initial model	Final model		Initial model	Final model
HDK1	Correct	Correct	LDK1	Incorrect	Not stated
HDK2	Incorrect	Incorrect	LDK2	Incorrect	Incorrect
HDK3	Ambiguous	Correct	LDK3	Incorrect	Incorrect
HDK4	Correct	Correct	LDK4	Incorrect	Not stated
HDK5	Ambiguous	Correct	LDK5	Incorrect	Incorrect
HDK6	Incorrect	Incorrect	LDK6	Incorrect	Incorrect
HDK7	Ambiguous	Integrated[a]	LDK7	Incorrect	Incorrect
HDK8	Correct	Correct	LDK8	Incorrect	Not stated
HDK9	Incorrect	Correct	LDK9	Incorrect	Incorrect
HDK10	Correct	Correct			

[a] This participant reported both air entering the siphon bell and the disks closing as explanations for the end of the outlet process.

5.2.4. Troubleshooting

The HDK participants on average generated significantly more troubleshooting responses ($M = 8.50$, $\mathrm{SD} = 2.22$) than the LDK participants ($M = 6.22$, $\mathrm{SD} = 2.49$, $F(1, 17) = 4.44$, $p = .05$, $\eta^2 = .21$). This difference in means is shown in Fig. 5.

5.3. Discussion

In examining how prior knowledge affects learning from realistic animations without accompanying verbal instruction, the data suggest that participants with a background in physics and engineering are more able to construct an appropriate understanding of a mechanical system from an animation than their LDK counterparts. The HDK participants were able to extract more steps from the animation initially, were more accurate in answering an explicit question about how the outlet system functions, and, upon a second viewing of the materials, were more likely to revise their incorrect models of the output system. Furthermore, they were able to generate more plausible reasons for system breakdown in the troubleshooting questions. The LDK group, on the other hand, reported incorrect models of the outlet system and did not revise these models even when they described the mechanical system while viewing the animations.

One surprising result of this experiment was that some HDK participants were not initially successful at building the correct model of the flushing cistern. In fact, only four of the 10 participants reported the correct mental model and three reported the disks stopping the water in answering the function questions, suggesting that they too were susceptible to constructing an erroneous mental model. This result is consistent with much of the naïve physics literature (e.g., Clement, 1983; McCloskey, 1983), which shows that naïve misconceptions often remain after students have acquired formal knowledge in physics. It appears that the erroneous mental model was initially in competition with the physically accurate explanation for the HDK participants, even though they had formal knowledge of mechanics. A novel result of this study was that after an additional learning iteration, most HDK learners were able to critique their original explanations and arrive at the correct explanatory schema.

These results provide some insight into the process of constructing understanding from animations. Most interesting are the patterns of the HDK participants. In the initial step-by-step descriptions, only two of these 10 participants stated the correct model of the output system. Later, when specifically asked the question of how the outlet process ends, four stated the correct answer. Finally, when orally reporting the "events" of the system, seven of the 10 mentioned the process ending correctly. It seems that being explicitly asked how the water stops flowing, as well as viewing the materials for a second time caused many of the HDK participants to revise their incorrect models into physically correct models. Regardless of whether this is due to a self-explanation effect (Chi et al., 1994) or a second exposure to the learning materials, it is important to note that none of the participants were ever provided with additional information about the system.

As previous literature has suggested, learning can be an iterative process of misunderstanding and understanding (Miyake, 1986; Rozenblit and Keil, 2002). In this experiment, which evaluated an extended learning session, the iterative nature of the learning session seemed to benefit only those who had previously acquired adequate domain knowledge. Iterative learning sequences provide no additional benefit to LDK participants, perhaps because they do not possess the knowledge necessary to perceive informational gaps in their models (Chi, 2000). This study indicates that some of the learning patterns documented in classic learning studies (i.e., Chi et al., 1994; Miyake, 1986) can also be found in iterative learning with animation.

6. General discussion

The findings presented in this paper indicate that animations designed to alleviate difficulties in processing animations, due to their transience and their demands on visual attention, do not necessarily contribute to a better understanding of mechanical devices. Neither interactivity nor signaling helped participants construct a correct mental model of a mechanical system in Experiments 1–3. Furthermore, participants who received interactive animations and animations with signaling devices typically developed a mental model of the system that was inconsistent with what was shown in the animation. On the other hand, the results of Experiment 4 suggest that prior knowledge contributes to constructing understanding from these materials. These results suggest that optimizing learning from animations is not merely a question of improving the external display and that researchers should put more focus on how to best prepare students for learning from animations.

A unique methodological contribution in this study was the examination of the specific mental models that students built. To complement the quantitative measures of learning provided by the comprehension questions, analysis of the in-depth explanations that students provided allowed us to evaluate how learners understood the mechanical system as a cohesive whole. This was especially important in revealing consistent misconceptions about how the system worked as well as situations in which participants constructed the correct mental model from the animations. Evaluating mental models also enabled us to examine how knowledge structures changed (or did not change) with additional exposures to the materials.

Analysis of students' correct and incorrect mental models suggest that prior knowledge, whether domain general or domain specific, was applied to the learning materials. Accurate mental models of the inlet and outlet systems depend on understanding of two "invisible" forces; buoyancy and siphoning. In the inlet system, the buoyant properties of the float allow it to move up and down with the water level to regulate the flow of water through the inlet valve. The results of all four experiments indicate that learners understand this mechanism well. On the other hand, the LDK learners, and even some of the HDK learners, had much more difficulty understanding the siphon process that occurs in the outlet system and exhibited a consistent misconception that the disks stopped the outflow of water.

The crucial difference between these two processes may be the accessibility of analogous explanation from everyday knowledge. Researchers have proposed that over the course of a lifetime, one's perceptual experiences in everyday life become organized into an explanatory framework, referred to as naïve physics understanding (e.g., Clement, 1983; McCloskey, 1983; diSessa, 1993; Kozhevnikov and Hegarty, 2001; Vosniadou, 2002). Most adults have accumulated a number of "everyday" real-world experiences regarding the phenomenon of floating. Thus, naïve physics may provide an adequate knowledge base to interpret the function of the float. It should also be noted that the same type of float is used to control the inlet mechanism in many American toilets, so participants were also more likely to have specific knowledge of this mechanism.

In contrast, siphoning is a less common everyday phenomenon, and the common presupposition found in most participants' descriptions was that solid objects (e.g., the disks) stop fluids (e.g., water). This may be due to a naïve physics explanation based on everyday experiences with the principle of damming. Additionally, participants could have tried to transfer their real-world knowledge of American toilets to their understanding of these visual materials. Without access to an explanation that better matched the steps of the animations, the LDK learners may have interpreted the disks' function like the rubber flap of an American toilet.

An important conclusion of this study is that attending to the correct locations in an animation or having control over the pace of the animation are not sufficient to fill informational gaps caused by an inadequate knowledge base. For instance, if participants do not have the background knowledge necessary to understand the dynamic process of siphoning, a signaling device pointing to the end of the siphoning process or the ability to slow the animation during this step does not provide explicit information about why and how this process occurs. Indeed, Experiment 3 indicated that participants spent more time examining the end of the siphoning process and signaling devices were successful in increasing eye fixations to the relevant location on the screen, yet participants failed to understand this process.

This study focused on comprehension of animations presented alone without verbal instruction. Many previous studies have shown that both animations and static illustrations are more effective when they are accompanied by text or verbal commentaries than when they are presented alone (e.g., Mayer and Gallini, 1990; Hegarty and Just, 1993; Mayer, 2001; Hegarty et al., 2003) and this is especially true in the case of novices (Kalyuga et al., 1998). Oestermeier and Hesse (2000) have suggested that in contrast to diagrams and animations, text or verbal commentaries are capable of expressing both abstract and concrete ideas equally, so that verbal codes integrated with static diagrams or dynamic visualizations can convey information that may be less easily communicated through visuo-spatial materials alone. An accompanying text can also be powerful in directing a learner's attention to elements in the display that might otherwise have been missed (cf. Hegarty and Just, 1993). This raises the question of whether LDK individuals might construct more accurate mental models of the siphon process from animations accompanied by verbal descriptions.

In our own previous research (Hegarty et al., 2003), we collected data on learning about the flushing cistern from a variety of instructional materials including animations with accompanying verbal commentaries. To examine how verbal information may aid learning from animation, we re-analyzed the data from that study to assess whether participants constructed the correct mental model of the outlet system when they were given an animation with a simultaneous audio commentary of how the flushing cistern works. In their step-by-step reports of the causal chain, 36% of these participants reported the correct model of how the outlet process ended. This result is in striking contrast with the data collected in the present study, in which only three, or 2.1% of the LDK participants across four experiments correctly reported the ending of the outlet process from the animations alone. Thus, giving LDK learners explicit information about the physical processes in the form of an accompanying verbal description may lead to more accurate mental models.

Although we recommend the addition of verbal information to enhance students' learning from animations, there are two reasons why designing instructional materials that provide explicit, domain-specific information in an accompanying text may not be a sufficient solution for all students. First, our research suggests that learners often seem to ignore information in a display that contradicts their prior intuitions about the material, so that merely telling them or showing them how a mechanism works may not be enough to overcome prior misconceptions for all learners. Note that only 36% of the students in Hegarty et al.'s (2003) experiment constructed the correct mental model from the animation and commentary. Second, while it may seem that extra information cannot hurt, many studies have shown what Kalyuga et al. (2003) call the "expertise reversal effect." Their review of empirical work on the interaction between domain knowledge and learning suggests that techniques that are highly effective for novice learners can have negative outcomes if used on expert learners. For learners who have prior domain knowledge, iterative learning experiences with non-verbal animations may be more effective for building a complete understanding of the system-at-hand. Importantly, there is probably no universal formula for designing a good animation, because the most effective animation will depend on characteristics of the user and the context in which it is used.

From a methodological perspective, most of the research to date on learning from animations has given students just a single learning experience, and it is important to study the effectiveness of animations in more iterative learning situations that are more characteristic of real-world learning situations, as we did in Experiment 4. It is striking that even some HDK individuals in that experiment needed multiple iterations of learning to construct the correct mental models. This suggests that rather than a "one shot" approach, a more effective use of animations may be to embed them in a curriculum in which students are induced to make their mental models explicit and view the animations in several iterations.

Although interactivity and cueing did not benefit the LDK learners in the mechanical domain used this study, this does not mean that these design features of animations can never provide any benefit. For example, Grant and Spivey (2003) found that animating the healthy tissue in a picture illustrating Duncker's tumor problem caused more participants to reach the correct solution to this insight problem (an effect of attentional signaling). Furthermore, Schwan and Riempp (2004) presented two groups of people learning to tie nautical knots with two non-verbal videos—one was interactive and the other was not. They found that learners using the interactive video benefited more than those who learned from a video that played at a steady rate (an effect of interactivity).

Why, then, are design features like signaling and interactivity effective in some cases and not in others? One common factor in the previous studies is the participants' reliance on domain general knowledge. Insight problems can be solved using creative thinking, but the solutions do not require specialized knowledge, so all participants in Grant and Spivey's (2003) study had relevant prior knowledge, that in addition to the signaling, could guide them to the correct answer. Similarly, in Schwan and Riempp's (2004) study, the process of viewing and copying the motions in the videos was presumably based on general knowledge. These results suggest that when participants have the necessary knowledge, signaling and interactivity can be effective. However, when people do not have necessary prior knowledge, as in our study, or Lowe's (1999, 2003) research on meteorology, they are less able to benefit from enhanced animations. Complexity of the dynamic information may also dictate the effectiveness of design features (Schwan and Riempp, 2004). Compared to knot-tying or Duncker's tumor problem, the flushing cistern and meteorological patterns involve many more elements and relations that must be integrated into a cohesive mental model. Thus, signaling and interactivity may be effective when the information to be provided involves a relatively low number of elements and relations. Of course these explanations are speculative at this stage, and more research must be conducted to understand the conditions under which signaling and interactivity help learners.

A possible limitation of this study is that all of the experiments examined learning of the same mechanical system. Naturally, one wonders how well our results generalize to other multimedia materials. In another recent study that was concerned with learning from animated versus static materials (Mayer et al., 2005), results were consistent across a variety of different animations, including the flushing cistern animation used in this research and animations of brakes, ocean waves, and lightning. Thus we expect the results from this experiment to generalize.

In conclusion, educators and multimedia designers should consider that problems concerning learning from animations cannot always be solved simply by using general design principles to build "better" animations without taking the learning context and learners' prior knowledge or level of expertise into account. As we continue to explore how to use this new technology in education, it will be important to consider both the top-down and bottom-up influences on learning from animations and other types of external visualizations and to take characteristics of the user, the learning goal, and the instructional context into consideration in designing instructional animations.

Acknowledgment

The authors would like to thank Jerry Tietz, John Nychka, and Kevin Beer for lending their artistic talents and programming expertise, as well as Marriette Riad, Farbod Bayegan, and Autumn Elling for their help with data collection. This research was supported by the Office of Naval Research Grant no. N00014-03-1-0119 to Mary Hegarty.

References

Ainsworth, S., Van Labeke, N., 2004. Multiple forms of dynamic representation. Learning and Instruction 14, 214–255.

ChanLin, L., 1998. Students' cognitive styles and the need of visual control in animation. Journal of Educational Computing Research 19 (4), 353–385.

Chi, M., 2000. Self-explaining expository texts: the dual process of generating inferences and repairing mental models. In: Glaser, R. (Ed.), Advances in Instructional Psychology, vol 5: Educational Design and Cognitive Science. Lawrence Erlbaum, Mahwah, NJ, pp. 161–238.

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

Clement, J., 1983. A conceptual model discussed by Galileo and used intuitively by physics students. In: Gentner, D., Stevens, A. (Eds.), Mental Models. Erlbaum, Hillsdale, NJ, pp. 325–339.

diSessa, A., 1993. Toward an epistemology of physics. Cognition and Instruction 10, 105–225.

Faraday, P., Sutcliffe, A.G., 1997a. An empirical study of attending and comprehending multimedia presentations. Proceedings of ACM Multimedia 96 Conference, 265–275.

Ferguson, E.S., 1992. Engineering and the Mind's Eye. MIT Press, Cambridge, MA.

Gentner, D., Stevens, A.L., 1983. Mental Models. Erlbaum, Hillsdale, NJ.

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

Grant, E.R., Spivey, M.J., 2003. Eye movements and problem solving: guiding attention guides thought. Psychological Science 14 (5), 462–466.

Hegarty, M., 2004. Dynamic visualizations and learning: getting to the difficult questions. Learning and Instruction 14, 343–351.

Hegarty, M., 2005. Multimedia learning about physical systems. In: Mayer, R.E. (Ed.), The Cambridge Handbook of Multimedia Learning. Cambridge University Press, Cambridge, pp. 447–465.

Hegarty, M., Just, M.A., 1993. Constructing mental models of machines from text and diagrams. Journal of Memory and Language 32, 717–742.

Hegarty, M., Kriz, S., Cate, C., 2003. The roles of mental animations and external animations in understanding mechanical systems. Cognition and Instruction 21 (4), 325–360.

Kalyuga, S., Chandler, P., Sweller, J., 1998. Levels of expertise and instructional design. Human Factors 40 (1), 1–17.

Kalyuga, S., Ayres, P., Chandler, P., Sweller, J., 2003. The expertise reversal effect. Educational Psychologist 38, 23–31.

Kintsch, W., 1988. The role of knowledge in discourse comprehension: a construction integration model. Psychological Review 95, 163–182.

Kozhevnikov, M., Hegarty, M., 2001. Impetus beliefs as default heuristics: dissociation between explicit and implicit knowledge about motion. Psychonomic Bulletin & Review 8, 439–453.

Larkin, J.H., Simon, H.A., 1987. Why a diagram is (sometimes) worth 10,000 words. Cognitive Science 11, 65–99.

Lowe, R., 1994. Selectivity in diagrams: reading beyond the lines. Educational Psychology 14 (4), 467–491.

Lowe, R.K., 1999. Extracting information from an animation during complex visual learning. European Journal of Psychology of Education 14 (2), 225–244.

Lowe, R.K., 2003. Animation and learning: selective processing of information in dynamic graphics. Learning and Instruction 13, 157–176.

Lowe, R.K., 2004. Interrogation of a dynamic visualization during learning. Learning and Instruction 14, 257–274.

Macaulay, D., 1998. The New Way Things Work. CD-ROM. DK Interactive Learning, New York, NY.

Mayer, R.E., 2001. Multimedia Learning. Cambridge University Press, New York.

Mayer, R.E., Anderson, R.B., 1991. Animations need narrations: an experimental test of the dual-code hypothesis. Journal of Educational Psychology 83, 484–490.

Mayer, R.E., Gallini, J., 1990. When is an illustration worth ten thousand words? Journal of Educational Psychology 83, 715–726.

Mayer, R.E., Moreno, R., 2003. Nine ways to reduce cognitive load in multimedia learning. Educational Psychologist 38 (1), 43–52.

Mayer, R.E., Hegarty, M., Mayer, S., Campbell, J., 2005. When static media promote active learning: diagrams versus animations in multimedia instruction. Journal of Experimental Psychology 11, 256–265.

McCloskey, M., 1983. Naïve theories of motion. In: Gentner, D., Stevens, A. (Eds.), Mental Models. Erlbaum, Hillsdale, NJ, pp. 229–324.

Miyake, N., 1986. Constructive interaction and the iterative process of understanding. Cognitive Science 10, 151–177.

Narayanan, N.H., Hegarty, M., 1998. On designing comprehensible interactive hypermedia manuals. International Journal of Human–Computer Studies 48, 267–301.

Narayanan, N.H., Hegarty, M., 2002. Multimedia design for communication of dynamic information. International Journal of Human–Computer Studies 57, 279–315.

Oestermeier, U., Hesse, F.W., 2000. Verbal and visual causal arguments. Cognition 75, 65–104.

Pane, J.F., Corbett, A.T., John, B.E., 1996. Assessing dynamics in computer-based instruction. In: Tauber, M.J. (Ed.), Proceedings of the ACM Conference on Human Factors in Computing Systems. ACM, Vancouver, pp. 797–804.

Rozenblit, L.G., Keil, F.C., 2002. The misunderstood limits of folk science: an illusion of explanatory depth. Cognitive Science 26, 521–562.

Schwan, S., Riempp, R., 2004. The cognitive benefits of interactive videos: learning to tie nautical knots. Learning and Instruction 14, 293–305.

Smallman, H.S., St. John, M., 2005. Naïve realism: misplaced faith in realistic displays. Ergonomics in Design 13, 14–19.

Tversky, B., Morrison, J.B., Betrancourt, M., 2002. Animation: can it facilitate? International Journal of Human–Computer Studies 57, 247–262.

Vosniadou, S., 2002. On the nature of naïve physics. In: Limon, M., Mason, L. (Eds.), Reconsidering Conceptual Change: Issues in Theory and Practice. Kluwer, Dordrecht, pp. 61–76.