Learning With Animation and Illusions of Understanding

Eugene S. Paik and Gregory Schraw
University of Nevada, Las Vegas

The illusion of understanding hypothesis asserts that, when people are learning with multimedia presentations, the addition of animation can affect metacognitive monitoring such that they perceive the presentation to be easier to understand and develop more optimistic metacomprehension. As a result, learners invest less cognitive effort when learning with animation. This study tested the illusion of understanding hypothesis with a randomized, double-blind, 2 × 2 factorial design using two different types of animation—representational and directive. Representational animation had a negative effect on learning, and directive animation had a positive effect. Both representational and directive animations induced illusion of understanding. Moreover, the animations induced multiple forms of the illusion. Consistent with expertise reversal effect, the animations induced more optimistic metacomprehension in low-proficiency learners but more pessimistic metacomprehension in high-proficiency learners.

Keywords: animation, multimedia, metacognition, metacomprehension, expertise reversal effect

Supplemental materials: http://dx.doi.org/10.1037/a0030281.supp

Research shows that adding animation to multimedia presentations can have negative as well as positive effects on learning (Betrancourt, 2005; Betrancourt & Tversky, 2000; Höfller & Leutner, 2007; Tversky & Morrison, 2002). Multimedia theorists have postulated that animation can impede learning by perturbing metacognitive monitoring. In particular, animation can induce an illusion of understanding in which learners develop more optimistic metacomprehension. As a result, learners reduce their cognitive engagement when learning with animation (Betrancourt, 2005; Kühl, Scheiter, Gerjets, & Gemballa, 2011; Lewalter, 2003; Schnotz & Rasch, 2005).

Numerous studies have examined the effects of animation on learning, but few have systematically examined the effects of animation on metacognitive processes during learning. The present study fills this gap by examining claims made by the illusion of understanding hypothesis and the accuracy of performance standard hypothesis. This examination was conducted with two different types of animation—representational and directive. The results promote a clearer theoretical understanding of how animation affects learning through its influence on metacognition.

**Representational and Directive Animation**

Representational animation and directive animation embody two distinct cognitive strategies for helping the learner. An animation is said to be representational if it illustrates the content of the presentation (Höfller & Leutner, 2007; Paik, 2009; Schnotz & Lowe, 2008). Typically, representational animations have been used to depict the behavior of dynamic systems as they change over time. For example, in this study representational animation was used to show the movement of the parts in a flushing toilet tank (see Figure 1). Although representational animations can explicitly portray the dynamic behavior of the system, such behaviors can only be implied with static images (Betrancourt, 2005; Schnotz & Lowe, 2003, 2008). Therefore, with static images, the learner must infer the system’s behavior. In the extreme case, representational animation can enable the learner to form a mental visualization of the system’s behavior that the learner is incapable of forming with only static images. Even when the learner is capable of inferring the system’s behavior from static images, representational animation can facilitate the inferential process and, thereby, reduce cognitive demand (Schnotz & Lowe, 2008; Schnotz & Rasch, 2005).

An animation is said to be directive when it directs the viewer’s attention to a particular component or area of an image (Schnotz & Lowe, 2008). The concept of directive animation has had numerous labels in the literature including highlighting (Jeung, Chandler & Sweller, 1997), signaling (Kriz & Hegarty, 2007), and cuing (de Koning, Tabbers, Rikers, & Paas, 2009). Directive animations can be implemented with a variety of techniques including flashing (i.e., quickly lightening and darkening an image area) or displaying an image area in different colors, as well as with animated pedagogical agents (e.g., cartoon characters) that point to or look in the direction of an image area (de Koning et al., 2009; Mayer, 2005; Moreno, 2005). Directive animations typically have been used to help the learner integrate aural and visual components of multimedia presentations. By synchronizing directive animation with running narration, directive animation can support the process of searching the image for the referents of the narration. In this manner, directive animation can increase the cognitive efficiency of visually identifying thematically salient components of an im-
Figure 1. Six static images corresponding to the six segments of multimedia presentation on the working of a flushing toilet tank.
age, thereby enhancing learning (Jeung et al., 1997; Schnott & Lowe, 2008).

**Illusion of Understanding**

The addition of directive animation in multimedia presentations generally has a positive effect on learning when the animation helps the learner integrate aural and visual components of the presentation (Höffler & Leutner, 2007). In contrast, the addition of representational animation in multimedia presentations can have negative as well as positive effects on learning (Betrancourt, 2005; Betrancourt & Tversky, 2000; Höfler & Leutner, 2007; Tversky & Morrison, 2002). Multimedia theorists have proposed a number of cognitive mechanisms by which representational animation may impede learning. For example, Hegarty, Kriz, and Cate (2003) provided some support for the idea that cognitive representation (i.e., the manner in which the behavior of dynamic systems is represented in the cognitive system) is more similar to series of static images than to animations. Mayer, Hegarty, Mayer, and Campbell (2005) provided a plausible rationale for why germane cognitive processes are engendered more by static images than by representational animations. Others argued that representational animation impedes learning because animations can negatively impact metacognition. That is, representational animations can induce an illusion of understanding in learners (Betrancourt, 2005; Kühl et al., 2011; Lewalter, 2003; Schnott & Rasch, 2005).

According to the illusion of understanding (IU) model, adding animation to multimedia presentations can cause learners to overestimate how easy it is to comprehend the presented material and, thereby, develop inflated metacomprehension. As a result, learners invest less cognitive resources to the learning task. This effect is similar to the illusion of knowing phenomenon described in calibration research (Glenberg, Wilkinson, & Epstein, 1982; Serra & Metcalfe, 2009) in which learners overestimate the degree to which they understand information presented in text. As a result, learners allocate less attention and fail to monitor with full vigilance due to inappropriate judgments of learning.

We define the IU hypothesis as three metacognitive assertions. These assertions are based on a view of self-regulation as a dynamic cycle in which monitoring informs control processes that facilitate ongoing planning and implementation of strategies and subsequent monitoring (Azevedo & Witherspoon, 2009; Efklides, 2008; Nelson & Narens, 1990; Terricone, 2011; Winne, 2001). The judgment of difficulty (JOD) assertion states that, when people are learning about the behavior of dynamic systems with multimedia presentations, the addition of representational animation that explicitly illustrates the system’s behavior causes them to perceive the presentation as being easier to learn (Betrancourt, 2005; Salomon, 1984; Schnott & Rasch, 2005, 2008). With representational animation, mentally visualizing the behavior of the system is essentially a perceptual exercise. In contrast, when the behavior of the system is implied with static images, the learner must expend cognitive resources to mentally animate the system.

The judgment of comprehension (JOC) assertion states that the addition of representational animation causes learners to form more optimistic metacomprehension (Schnott & Rasch, 2005). The relative ease with which learners visually experience the system’s dynamic behavior with representational animation causes them to inflate their judgment of how well they comprehend that system. Serra and Metcalfe (2009) used the term fluency heuristics to refer to an analogous relationship reported in text comprehension research in which metacognitive judgments of task difficulty (e.g., retrieval fluency, ease of learning) influence judgments of comprehension. Also, Serra and Dunlosky (2010) reported that beliefs about the efficacy of multimedia formats influence learners’ metacomprehension judgments.

The disengagement assertion states that the addition of representational animation causes learners to reduce their cognitive engagement (Schnott & Rasch, 2005). An inflated sense of comprehension leads learners to conclude that less cognitive resources and effort are needed to sufficiently comprehend the presented information.

**Prior Studies**

Two think-aloud studies (Kühl et al., 2011; Lewalter, 2003) provided some support for the IU hypothesis based on the frequency of comprehension and planning statements. First, in both studies animation learners uttered more positive comprehension statements than static image learners. Consistent with the JOC assertion, this difference may have resulted from animation learners having developed an inflated sense of comprehension relative to that of static image learners. Second, in the Lewalter study, animation learners uttered fewer planning statements than static image learners. Consistent with the disengagement assertion, this difference may indicate that animation learners invested less cognitive effort than static image learners in remediation.

Although the two findings described above may be consistent with animation learners having experienced an illusion of understanding, the overall pattern of data from Lewalter (2003) and Kühl et al. (2011) is also consistent with an alternate explanation, that the differences in the frequency of comprehension and planning statements were caused by differences in learning. Because the differences in learning between static image learners and animation learners were statistically not significant in the two studies, there is the possibility that animation had a positive effect on learning, but this effect was undetected due to lack of statistical power. If so, increased learning by animation learners may have been responsible for their higher frequency of comprehension statements and lower frequency of planning statements. Two additional findings in Kühl et al.’s study cast further doubt that their animation learners experienced an illusion of understanding. First, Kühl et al.’s animation learners uttered fewer erroneous statements than did static image learners, indicating that animation may have improved comprehension. Second, no significant differences in judgment of difficulty or mental effort were detected between animation and static image learners. Taken together, the results of the Lewalter study and the Kühl et al. study do not provide conclusive evidence that animation had induced illusion of understanding.

**Accuracy of Performance Standard**

The IU hypothesis explicates a metacognitive process by which representational animation may impede learning. We now introduce a metacognitive process by which representational animation may enhance learning. We generated the accuracy of performance standard (APS) hypothesis because it constitutes a plausible and theoretically important counterargument to the IU hypothesis.
According to the APS hypothesis, when the behavior of a dynamic system is explicitly illustrated with representational animation, the learner has access to a more reliable standard in memory (i.e., mental representation) by which to evaluate his or her comprehension. In contrast, when the behavior of dynamic systems is presented as a set of static images, the learner’s meta-comprehension must depend on the behavior of the system that the learner infers. These inferences may be inaccurate or incomplete. Therefore, the APS hypothesis asserts that, when people are learning about the behavior of dynamic systems with multimedia presentations, the addition of representational animation that explicitly illustrates the system’s behavior causes them to generate more accurate JOC. This accuracy assertion has significant ramifications for the efficiency of the self-regulatory system because learners are better able to plan, select strategies, and allocate resources (Azevedo & Witherspoon, 2009; Nelson & Narens, 1990; Terricone, 2011; Tobias & Eveson, 2009; Winne, 2001). Serra and Metcalfe (2009) referred to this sequence as the accuracy–control link by which accurate monitoring enhances control, which enhances self-regulation.

**Present Study**

This study tested the assertions of the IU and the APS hypothesis with both representational animation and directive animation. The experiment had four treatment conditions: static (i.e., no animation); representational animation only; directive animation only; and both representational and directive animations. Following the precedent of Mayer et al. (2005), two types of learning outcomes were measured. A retention test asked the participants to recall information explicitly provided in a multimedia presentation. A transfer test asked the participants to solve diagnostic and prognostic problems.

Judgment of difficulty (JOD), judgment of comprehension (JOC), and judgment of visualization (JOV) were used as dependent variables to test the assertions of the IU and APS hypotheses. JOD was estimated by asking the participants to characterize their learning experience (e.g., “How difficult was it to learn about the flushing toilet tank from the presentation?”) and the multimedia presentation (e.g., “How would you rate the quality of the presentation that you just saw?”). JOC was estimated by asking the participants to predict their learning outcome. After the treatment was administered, the participants were asked to estimate how they would perform on a number of different problems. As these were the problems in the instruments used to measure their learning outcome (i.e., the retention test and the transfer test), the participants were effectively predicting their learning outcome. JOV was measured by asking the participants to mentally visualize the behavior of the flushing toilet tank that they had learned about during the treatment. They were then asked to characterize the quality of their visualization experience in terms of difficulty, accuracy, and level of detail.

The analysis utilized two covariates: spatial ability and prior knowledge. Prior studies showed that spatial ability and image type (e.g., animation vs. statistic images) can have interactional effects on learning (Höffler & Leutner, 2007). Prior knowledge had been shown to affect learning in general (Kalyuga, Ayres, Chandler, & Sweller, 2003).

**Predictions**

We did not include predictions of learning outcome because a number of potential cognitive mechanisms have been hypothesized by which animation may influence learning. As yet, however, there is an absence of a unifying framework that describes (a) the conditions under which each of these mechanisms become activated and (b) how the activated processes are integrated to produce the learning outcome. Indeed, one of our goals for this study was to contribute to the theoretical advancement with regard to (a).

The JOD assertion of the IU hypothesis was tested with the null hypothesis that representational animation learners would produce lower JOD (i.e., find the presentation easier) than no representational animation learners. The JOC assertion of the IU hypothesis was tested with the null hypothesis that representational animation learners would produce higher JOC (i.e., form more optimistic metacomprehension) than no representational animation learners. The disengagement assertion of the IU hypothesis was tested tangentially with JOV. According to the disengagement assertion, representational animation learners invest less cognitive resources and effort than no representational animation learners when learning about the behavior of a dynamic system. One cognitive activity that representational animation learners are less likely to engage in is mentally simulating the behavior of the system (Schnotz & Rasch, 2005). If so, then representational animation learners will have more difficulty mentally visualizing the system after the learning phase because they had less practice mentally visualizing the system during learning. Hence, the disengagement assertion was tested with the null hypothesis that representational animation learners would produce lower JOV (i.e., poorer quality of visualization) than no representational animation learners. The accuracy assertion of the APS hypothesis was tested with the null hypothesis that representational animation learners are less likely to engage in and mentally visualize the behavior of the flushing toilet tank from the presentation?"). JOC was estimated by asking the participants to predict their learning outcome. After the treatment was administered, the participants were asked to estimate how they would perform on a number of different problems. As these were the problems in the instruments used to measure their learning outcome (i.e., the retention test and the transfer test), the participants were effectively predicting their learning outcome. JOV was measured by asking the participants to mentally visualize the behavior of the flushing toilet tank that they had learned about during the treatment. They were then asked to characterize the quality of their visualization experience in terms of difficulty, accuracy, and level of detail.

The analysis utilized two covariates: spatial ability and prior knowledge. Prior studies showed that spatial ability and image type (e.g., animation vs. statistic images) can have interactional effects on learning (Höffler & Leutner, 2007). Prior knowledge had been shown to affect learning in general (Kalyuga, Ayres, Chandler, & Sweller, 2003).

**Analytical Methods**

Groupwise comparisons of JOD, JOC, and JOV were made not with raw scores but with bias scores. The bias of a variable was defined as the component of the variable that is not attributable to learning. The bias of a variable was estimated with the residual when the variable is regressed by overall learning (i.e., the average of retention and transfer test scores). This analytical approach eliminated the potential explanation that a detected difference in a variable was caused by differences in learning.
The accuracy of JOC for a group was estimated with the Pearson correlation between JOC and overall learning (Blanch-Hartigan, 2011). Groupwise comparisons of JOC accuracy were made with Fisher’s Z-transformation.

### Method

This study used a randomized, double-blind, 2 × 2 between-subjects factorial design. The two factors were the absence or presence of representational animation and the absence or presence of directive animation. This resulted in four treatment groups: static (i.e., no animation), representational animation only, directive animation only, and both animations (i.e., representational animation and directive animation).

### Participants

The participants were 65 undergraduate psychology students (49 female, 16 male, M<sub>age</sub> = 24.3 years, age range = 18 to 44 years) at a major university in the southwestern region of the United States. We do not report the participants’ SAT scores because a majority of the participants did not provide them.

### Material

The experiment protocol was administered by an interactive multimedia computer program that recorded all participant responses. All the participants worked on an identical model of computer hardware that included a 24-in. color monitor with 1920 × 1080 resolution and a headphone. A version of the computer program that was used to administer the protocol is provided as online supplemental material. With this computer program, the reader may examine the entire protocol including the instruments, the treatment (i.e., the four versions of the multimedia presentation on the workings of a flushing toilet tank), as well as the manner in which each instrument and treatment condition was administered. The following describes the components of the protocol in the order that they were administered.

#### Introduction

The introduction explained what was expected of the participants in the study. At the conclusion, the participants were asked to enter a treatment code that was provided to them by the protocol administrator.

#### Participant survey

The participant survey asked the participants about their demographic characteristics. The survey also asked the participants to characterize their background knowledge relevant to flushing toilet tanks.

#### Treatment

The treatment was a multimedia presentation on how a flushing toilet tank works that was adapted from Mayer et al. (2005) and Hegarty et al. (2003) with the addition that navigational control and the narration modality were controlled across the treatment conditions (Ginns, 2005; Low & Sweller, 2005; Mayer & Chandler, 2001; Mousavi, Low, & Sweller, 1995). Participants in this study viewed the presentation twice so as to provide a more realistic learning scenario than those in the Mayer et al. and Hegarty et al. studies.

The multimedia presentation contained six segments. Each segment contained graphic illustrations (see Figure 1) with an aural narration (see the Appendix). The first segment introduced the parts of the flushing toilet tank. The remaining segments described the five phases of generating the flush and refilling the tank. The first segment lasted about 60 seconds, and the remaining segments lasted about 30 seconds each. The participants viewed all the segments continuously from start to finish. The participants were not provided any mechanism (e.g., pause or rewind buttons) to interrupt the presentation. At the completion of the first viewing, the presentation was paused. When the participant clicked a button on the computer screen, the presentation of the second viewing commenced.

The images in the presentation contained a limited palette of colors. The parts of the toilet tank were in shades of gray. The water and the arrows representing the flow of water were in shades of blue. The arrows that indicated how the parts of the toilet tank moved were in red.

The multimedia presentations across the four treatment conditions were identical except for their visual component. The static version presented a single still image for each segment (see Figure 1). Each static image was a key frame from the animated versions. For several static images, however, water flow arrows were added to better illustrate the directionality of the water’s flow.

The directive-only version was identical to the static version except that several directive animation techniques were incorporated. In the first segment, parts appeared and disappeared from the screen in synchrony with the narration so that only those parts pertinent to the narration were displayed. For example, during the portion of the narration that stated “The flow of water into the tank is controlled by these parts . . .” only those enumerated parts were visible. Furthermore, the parts were displayed in a lighter shade until a part was explicitly referenced in the narration. At that point, the part being referenced was presented in normal shade, thereby providing the participant a clear visual indication of the part that was being referred to by the narration. In the remaining segments that described the dynamic behavior of the flushing toilet tank, blinking and tinting were used as visual cues to the narrative referents. For example, when the narration stated “When the handle is pressed down,” the color of the handle was tinted red and the arrow pointing down on the handle flashed on and off.

In the representational-only version, representational animation depicted the movement of the parts and the flow of water. The representational animation and the narration were synchronized. For example, when the narration in phase 2 stated that “The two disks start to drop and separate from each other,” the animation showed the two discs falling and separating from each other. The flow of water was animated by a continuous movement of arrows.

The representational–directive version incorporated both representational and directive animations described above. Logistically, we created first the representational–directive version such that directive animation and representational animation always appeared sequentially and never simultaneously. The other three versions of the presentation were then produced by replacing each unneeded animation clip with an appropriate static image.

#### Post-treatment survey

The post-treatment survey asked the participants to evaluate the format and the content of the multimedia presentation (e.g., “How difficult was it to learn about the flushing toilet tank from the presentation?”). The survey also described a set of problems to the participants. For each problem, the participants were asked how well they thought they would be able to solve the problem. For example, one of the questions was
as follows: “When a flushing toilet tank behaves abnormally, it is usually an indication that something in the tank is broken. How well do you think that you will be able to diagnose the cause of abnormal behaviors in flushing toilet tanks?” As the problem descriptions also described the problems used in the instruments to measure their learning outcome (i.e., the retention and transfer tests), the participants were asked, in effect, to predict their learning outcome.

Visualization exercise and survey. The visualization exercise asked the participants to mentally visualize the processes of the flushing toilet tank as described in the multimedia presentation. The participants were provided 30 seconds to perform the mental visualization. The visualization survey then asked the participants to characterize their visualization experience (e.g., “How detailed did your visualization seem to you? That is, compared to the level of detail provided in the presentation.”).

Retention test. The retention test required the participants to recall information that was explicitly provided in the presentation. There were 10 part-recall problems and one process-recall problem. For each part-recall problem, an image of a part of the flushing toilet tank was displayed on the screen. The image was identical to that shown during the treatment presentation. All of the other parts were also shown, but they were visually distinguished with a lighter shade. Each part-recall problem required the participant to provide two responses: part-name and part-purpose. The participants provided their responses by selecting an entry in two drop-down list boxes. For the part-name response, the list included names that were explicitly referred to in the presentation (e.g., connecting rod, float, float arm) as well as those that were not (e.g., regulator). Similarly, the list for the part-purpose response included concepts explicitly referred to in the presentation (e.g., flow of water into the tank) as well as those that were not (e.g., flow of air into the tank). The participants were given a maximum of 20 seconds to complete each part-recall problem. If a participant did not provide both responses within 20 seconds, the participant was notified that the time limit had been exceeded and the next question was automatically displayed.

The process-recall problem asked the participants to write down all the key events of the flushing toilet tank. The participants typed their responses into a basic text editor window. To help the participants understand what was expected of them, we provided three key events as a starting point: (a) the handle is pressed; (b) float drops toward the bottom of the tank; and (c) the inlet valve is pushed in the inlet pipe. The participants were provided a maximum of 3 minutes to respond.

Transfer test. The transfer test required the participants to apply the principles that were introduced in the treatment presentation in novel situations. The transfer test consisted of two prognostic problems (e.g., “Suppose that the float were to break off from the float arm. How would the flushing toilet tank misbehave? Describe all the symptoms that you can think of.”) and two diagnostic problems (e.g., “Suppose that you push down on the handle, but there is no flush. No water flows into the toilet bowl, none whatsoever. Describe all the causes that you can think of.”). The participants were provided 90 seconds to complete each prognostic problem and 120 seconds to complete each diagnostic problem.

Spatial ability test. The Paper Folding Test (Ekstrom, French, Harman, & Derman, 1976) was adapted for online administration. An earlier pilot study indicated that the Paper Folding Test posed significant cognitive demands on the participants. Therefore, only the first 10 problems of the original test were administered in the present study. In the original test, the first 10 problems were of comparable difficulty to the second 10. Also, the Paper Folding Test was administered at the end of the experiment to minimize the potential impact of cognitive fatigue induced by the test. Although this ordering left open the possibility that the treatment conditions systematically and differentially affected the participants’ performance of the Paper Folding Test, we judged this more advantageous than the other order (i.e., where the differential effects of the Paper Folding Test occurs prior to the treatment) because key statistical analyses of the study did not rely on the results of the Paper Folding Test (see Results). The participants were given a maximum of 20 seconds to complete each Paper Folding Test problem. If a participant did not respond within the 20 seconds, the participant was notified that the time limit had been exceeded and the next question was automatically displayed.

Procedure

The experiment was administered in four group sessions, ranging from 10 to 22 participants each, within a span of 1 week. Each participant was seated in front of a computer. The participants were instructed to put on the headphone and were shown how to adjust the volume. They were then instructed to begin their participation by pressing a button on the screen. Upon completion of the introductory presentation, each participant had the option of either terminating participation or signing the informed consent form and continuing participation. No participant chose to terminate participation.

When the participant submitted the signed consent form, the administrator gave the participant a sheet of paper with a treatment code. The participant was then informed that the administrator was no longer available for assistance. Participants were instructed to proceed as best they could if any issues or questions arose during the session. Once a participant entered the treatment code, a computer program administered the experiment protocol. At the completion of the protocol, the computer program instructed the participant to notify the administrator. When the participant notified the administrator, the administrator thanked the participant and directed the participant to leave the lab.

Random assignment. Random assignment and even distribution of the participants across the treatment groups were implemented as follows. The treatment codes that were distributed to the participants were ordered such that each consecutive four treatment codes (a) contained all four treatment conditions and (b) were shuffled in random order. The treatment codes were distributed in the order that the participants submitted their signed consent forms.

Double blind control. Double blind control of the experiment was achieved as follows. First, the sheets of paper containing the treatment codes were folded and stapled so that the codes were hidden. Therefore, when the administrator handed a treatment code paper to a participant, the administrator did not know to which treatment condition the participant was being assigned. Second, the administrator did not engage in any interaction, including eye contact, with a participant once the participant had opened the treatment code paper. Third, physical partitions were erected be-
tween the participants’ desks so that one participant could not see the computer screen of another participant. Finally, during the scoring process, participant responses were encoded and collated so that the scorers did not know to which individual or treatment group a response belonged.

**Variables and scoring.** In the descriptions that follow, “equally weighted sum” of scores indicates that the highest score in each subgroup was first scaled to 100%. The remaining scores were then scaled proportionately (i.e., without first scaling the lowest score to 0%). The following variables were employed in the present study:

- **Representational animation** indicated the absence or presence of representational animation in the treatment.
- **Directive animation** indicated the absence or presence of directive animation in the treatment.
- **Spatial ability** was the percentage of correct responses in the Paper Folding Test.
- **Prior knowledge** was the sum of the responses to questions 5 through 7 in the participant survey.
- **Judgment of difficulty (JOD)** was the sum of the responses to questions 1 through 4 in the post-treatment survey. Lower values signify greater ease and higher values signify greater difficulty. Note that although the reduction in mental effort is a key assertion of the IU hypothesis (i.e., the disengagement assertion), the survey item on mental effort (i.e., “How much mental effort was required to learn about the flushing toilet tank from the presentation?”) was aggregated into JOD due to challenges of developing a reliable instrument for mental effort.
- **Judgment of comprehension (JOC)** was the sum of the responses to questions 5 through 12 in the post-treatment survey.
- **Judgment of visualization (JOV)** was the sum of the responses to the three questions in the visualization survey.
- **Retention** was the equally weighted sum of the part-recall score and the process-recall score from the retention test. For the part-recall score, one point was awarded for each correct part-name response and one point was awarded for each correct part-purpose response. The procedure for scoring the process-recall problem was adapted from Hegarty et al. (2003). A key was developed that consisted of 24 distinct processes of the flushing toilet tank that were explicitly described in the presentation (e.g., “the connecting rod rises,” “the connecting rod pulls up the lower disk”). One point was awarded for each key process that the participant mentioned in his or her response. Fictitious or incorrect responses were ignored. The order of the processes was also ignored.
- **Transfer** was the equally weighted sum of the scores of the four problems in the transfer test. The transfer problems were scored based on a predefined key in a manner similar to the process-recall problem.
- **Overall learning** was the standardized equally weighted sum of retention and transfer.

**Results**

**Scores**

The five free-response problems (i.e., the process-recall problem in the retention test and the four problems in the transfer test) were scored by two individuals. The Pearson correlation between the two scorers was .87. All the discrepancies between the two scorers were resolved through discussion and consensus. JOC and JOV were transformed by $x^{3/2}$ to reduce their skewness. After the transformations, the Cronbach’s alpha for JOD, JOC, and JOV were .77, .88, and .87, respectively. Table 1 provides the descriptive statistics of the experimental variables. Table 2 provides the correlations between the variables.

**Learning Outcomes**

A $2 \times 2$ (representational animation × directive animation) factorial analysis of covariance (ANCOVA) was conducted using retention and transfer as dependent variables and spatial ability as a covariate. Prior knowledge was not used as a covariate because it was not significantly correlated with either dependent variable. Levene’s homogeneity of variance test was not significant in the analyses.

For retention, there were no significant main effects. Spatial ability had a positive effect on retention, $F(1, 64) = 18.4, p < .001, \eta^2 = .23$.

For transfer, main effects occurred with representational animation and with directive animation. Representational animation learners performed moderately worse on the transfer test than no-representational animation learners, $F(1, 64) = 5.38, p = .024, \eta^2 = .082$. Also, directive animation learners performed moderately better on the transfer test than no-directive animation learn-

### Table 1

**Descriptive Statistics of Experimental Variables by Treatment Condition**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Static M (SD)</th>
<th>Representative only M (SD)</th>
<th>Directive only M (SD)</th>
<th>Representative and directive M (SD)</th>
<th>Total M (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>17</td>
<td>16</td>
<td>16</td>
<td>16</td>
<td>65</td>
</tr>
<tr>
<td>Prior knowledge</td>
<td>10.4 (4.45)</td>
<td>12.1 (3.30)</td>
<td>9.94 (3.07)</td>
<td>9.19 (3.87)</td>
<td>10.4 (3.78)</td>
</tr>
<tr>
<td>Spatial ability</td>
<td>.571 (1.93)</td>
<td>.556 (.163)</td>
<td>.544 (.163)</td>
<td>.506 (.224)</td>
<td>.545 (.185)</td>
</tr>
<tr>
<td>Retention</td>
<td>.537 (.239)</td>
<td>.529 (.189)</td>
<td>.549 (.153)</td>
<td>.471 (.181)</td>
<td>.522 (.192)</td>
</tr>
<tr>
<td>Transfer</td>
<td>.421 (.166)</td>
<td>.359 (.184)</td>
<td>.531 (.181)</td>
<td>.394 (.131)</td>
<td>.426 (.175)</td>
</tr>
<tr>
<td>Overall learning</td>
<td>.48 (.188)</td>
<td>.44 (.16)</td>
<td>.54 (.14)</td>
<td>.43 (.15)</td>
<td>.47 (.16)</td>
</tr>
<tr>
<td>JOD</td>
<td>11.24 (7.16)</td>
<td>9.00 (3.86)</td>
<td>9.69 (3.72)</td>
<td>9.69 (4.14)</td>
<td>9.92 (4.93)</td>
</tr>
<tr>
<td>JOC</td>
<td>341.2 (146.3)</td>
<td>369.3 (99.1)</td>
<td>347.7 (73.9)</td>
<td>350.1 (78.0)</td>
<td>351.9 (102.3)</td>
</tr>
<tr>
<td>JOV</td>
<td>103.3 (29.5)</td>
<td>95.7 (30.2)</td>
<td>103.8 (20.1)</td>
<td>88.1 (30.5)</td>
<td>97.8 (28.0)</td>
</tr>
</tbody>
</table>

*Note. SD = standard deviation; JOD = judgment of difficulty; JOC = judgment of comprehension; JOV = judgment of visualization.*
ers, \( F(1, 64) = 4.64, p = .035, \eta^2 = .072 \). Representational animation \( \times \) directive animation interaction was not a significant predictor of retention or of transfer.

**Illusion of Understanding**

We regressed representational animation, overall learning, and representational animation \( \times \) overall learning (henceforth abbreviated as X \( R \times L \)) on dependent variables JOD, JOC, and JOV in three independent analyses. Also, the three regression analyses were replicated with directive animation in place of representational animation. These regression analyses provided a comparison of the dependent variable among participants with comparable learning, so that any detected difference in the dependent variable may not be attributed to variance in learning. When a significant interaction was found, we calculated the intersection point between the two regression lines (one for the absence and the other for the presence of respective animation). We then calculated the region of significance using the Johnson–Neyman method (Preacher, Curran, & Bauer, 2006). The results of these regression analyses are provided in Table 3. How these analyses support and refute of the assertions of the IU hypothesis is summarized in Table 4.

**Representational animation.** The effect of representational animation on JOD was consistent with the JOD assertion, and the effect of representational animation on JOC was consistent with the JOC assertion. However, the effects were consistent only for a subgroup of participants, namely, those with low overall learning scores. First, \( R \times L \) was a significant predictor of JOD and of JOC, indicating that the effects of representational animation on JOD and on JOC were moderated by overall learning. Second, for both JOD and JOC, the \( R \times L \) interaction was disordinal. That is, the regression lines intersected within the region of interest (i.e., overall learning within \( M \pm 2 SD \)). Third, the region of significance analysis revealed that, for overall learning less than \( -.070 SD \), representational animation learners developed lower JOD (i.e., more optimistic metacomprehension) than no-representational animation learners. JOV was not significantly affected by representational animation or by \( R \times L \). Therefore, the IU disengagement assertion was neither supported nor refuted by the analysis.

**Directive animation.** Contrary to the predictions of the IU hypothesis (i.e., directive animation would have no effect on JOD, JOC, and JOV), directive animation had a number of significant effects on JOD and on JOC. Further surprisingly, directive animation and representational animation had a strikingly similar pattern of effect on JOD, JOC, and JOV. First, the \( D \times L \) interaction was a significant predictor of JOD and of JOC. Second, for both JOD and JOC, the \( D \times L \) interaction was disordinal. Third, the lower bounds of the regions of significance were within the regions of interest (i.e., \( -.519 SD \) for JOD; \( -.84 SD \) for JOC). Finally, JOV was not significantly affected by directive animation or by \( D \times L \).

**Accuracy of Performance Standard**

The accuracy of JOC for a group was estimated with the correlation of JOC and overall learning (referred to below as the JOC/L correlation) within the group. The JOC accuracy of two groups was compared by comparing the JOC/L correlations of the two groups with Fisher’s Z-transformation.

The APS assertion (i.e., representational animation would produce more accurate JOC) was neither supported nor refuted by the analysis. For representational animation learners, the JOC/L correlation was \( r = .500, N = .32, p = .004 \). For no-representational animation learners, the correlation was \( r = .734, N = .33, p < .001 \). However, the difference between the two correlations was not statistically significant.

With respect to directive animation, the results contradicted the predictions of the APS hypothesis (i.e., directive animation would have no effect on JOC accuracy). Directive animation learners produced significantly less accurate JOC than no-directive animation learners. For directive animation learners, the JOC/L correlation was \( r = .395, N = .32, p = .025 \). For no-directive animation learners, the correlation was \( r = .730, N = .33, p < .001 \). The difference between the two correlations was significant, \( Z = 1.857, p = .05 \).

**Static Versus Animation Groups**

We also compared JOC accuracy between participants whose multimedia presentation did not contain any form of animation (i.e., the static image group) and those whose multimedia presentation contained some form of animation (i.e., representational-
Table 3

Regression Statistics for JOD, JOC, and JOV

<table>
<thead>
<tr>
<th>Dependent variable factor</th>
<th>β</th>
<th>B</th>
<th>LL</th>
<th>UL</th>
<th>t</th>
<th>p*</th>
<th>Intersection point</th>
<th>95% CI of B</th>
<th>Significance region</th>
</tr>
</thead>
<tbody>
<tr>
<td>Representational animation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>JOD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R</td>
<td>-.228</td>
<td>-.226</td>
<td>-4.250</td>
<td>-.202</td>
<td>-2.000</td>
<td>.032</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L</td>
<td>-.803</td>
<td>-3.956</td>
<td>-5.326</td>
<td>-2.586</td>
<td>-5.774</td>
<td>.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R × L</td>
<td>.374</td>
<td>2.786</td>
<td>.736</td>
<td>4.837</td>
<td>2.717</td>
<td>.009</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>JOC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R</td>
<td>.211</td>
<td>42.827</td>
<td>2.867</td>
<td>82.787</td>
<td>2.143</td>
<td>.036</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L</td>
<td>.804</td>
<td>82.322</td>
<td>55.269</td>
<td>109.375</td>
<td>6.085</td>
<td>.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R × L</td>
<td>-.228</td>
<td>-35.282</td>
<td>-75.766</td>
<td>5.202</td>
<td>-1.743</td>
<td>.086</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>JOV</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R</td>
<td>-.084</td>
<td>-4.677</td>
<td>-16.305</td>
<td>6.950</td>
<td>-0.804</td>
<td>.424</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L</td>
<td>-.563</td>
<td>15.785</td>
<td>7.913</td>
<td>23.657</td>
<td>4.010</td>
<td>.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R × L</td>
<td>.024</td>
<td>1.036</td>
<td>-10.744</td>
<td>12.816</td>
<td>0.176</td>
<td>.861</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Directive animation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>JOD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>-.015</td>
<td>-.143</td>
<td>-2.176</td>
<td>1.890</td>
<td>-.0141</td>
<td>.889</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L</td>
<td>-.764</td>
<td>-3.764</td>
<td>-5.113</td>
<td>-2.414</td>
<td>-5.578</td>
<td>.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D × L</td>
<td>.406</td>
<td>3.081</td>
<td>1.006</td>
<td>5.154</td>
<td>2.972</td>
<td>.004</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>JOC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>-.071</td>
<td>-14.353</td>
<td>-53.510</td>
<td>24.803</td>
<td>-0.733</td>
<td>.466</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L</td>
<td>.830</td>
<td>84.941</td>
<td>58.958</td>
<td>110.924</td>
<td>6.537</td>
<td>.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>JOV</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>-.110</td>
<td>-6.111</td>
<td>-17.418</td>
<td>5.196</td>
<td>-1.081</td>
<td>.284</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L</td>
<td>.657</td>
<td>18.402</td>
<td>10.900</td>
<td>25.905</td>
<td>4.905</td>
<td>.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D × L</td>
<td>-.078</td>
<td>-3.345</td>
<td>-14.873</td>
<td>8.183</td>
<td>-0.580</td>
<td>.564</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. CI = confidence interval; LL = lower limit; UL = upper limit; JOD = judgment of difficulty; JOC = judgment of comprehension; JOV = judgment of visualization; R = representational; D = directive; L = overall learning; broi = beyond region of interest.

*a .000 means p < .0005. b Regions below lower bound and above upper bound are significant at α = .05.

only, directive-only, or representational–directive). Animation learners produced significantly less accurate JOC than static image learners. For animation learners, the JOC/L correlation was r = .445, N = .48, p < .002. For static image learners, the correlation was r = .856, N = .17, p < .001. The difference between the two correlations was significant, Z = 2.613, p = .009.

Expertise Reversal Effect

The disordinal interactions of R × L and of D × L with respect to JOD and JOC indicated that the animations may exhibit expertise reversal effect. That is, the animations may have opposite effects on low- and high-proficiency learners. With representa-

Table 4

Confirmations and Refutations of IU and APS Assertions

<table>
<thead>
<tr>
<th>Hypothesis assertion</th>
<th>Measure</th>
<th>Representation animation</th>
<th>Prediction</th>
<th>Result</th>
<th>Directive animation</th>
<th>Prediction</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>IU hypothesis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>JOD assertion</td>
<td>JOD bias</td>
<td>R &lt; noR</td>
<td>Confirmed: L &lt; .070</td>
<td>D = noD</td>
<td>Refuted: D &lt; noD, where L &lt; -.519</td>
<td></td>
<td></td>
</tr>
<tr>
<td>JOC assertion</td>
<td>JOC bias</td>
<td>R &gt; noR</td>
<td>Confirmed: L &lt; .058</td>
<td>D = noD</td>
<td>Refuted: D &gt; noD, where L &gt; .580</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Disengagement assertion</td>
<td>JOV bias</td>
<td>R &lt; noR</td>
<td>n.s.</td>
<td>D = noD</td>
<td>n.s.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>APS hypothesis</td>
<td>JOC accuracy</td>
<td>R &gt; noR</td>
<td>n.s.</td>
<td>D = noD</td>
<td>Refuted: D &lt; noD</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. IU = illusion of understanding; APS = accuracy of performance standard; JOD = judgment of difficulty; JOC = judgment of comprehension; JOV = judgment of visualization; R = Presence of representational animation; noR = Absence of representational animation; D = Presence of directive animation; noD = Absence of directive animation; L = overall learning (the units of overall learning are standard deviations); n.s. = not significant.
tional animation, only the lower bounds of the regions of significance were within the regions of interest. Therefore, only one group of participants (i.e., the low-proficiency learners) produced statistically significant differences in JOD and JOC between representational animation and no-representational animation learners.

With directive animation, however, both the lower bounds and the upper bounds of the regions of significance were within the regions of interest. Therefore, two groups of participants (i.e., the low-proficiency learners and the high-proficiency learners) produced statistically significant differences in JOD and JOC between directive animation and no-directive animation learners. Consistent with the expertise reversal effect, the two groups experienced opposite forms of illusion of understanding. Low-proficiency learners experienced optimistic illusion of understanding, where animation learners find the presentation easier to understand and inflate their metacomprehension. High-proficiency learners experienced pessimistic illusion of understanding, where animation learners find the presentation more difficult to understand and deflate their metacomprehension.

Discussion

This study examined the effects of representational and directive animations on learning and metacognition. We tested the illusion of understanding (IU) hypothesis, which predicted that, when people are learning about the behavior of a dynamic system with multimedia presentations, the addition of representational animation that explicitly illustrates the system’s behavior would cause them to (a) find the presentations easier to understand, (b) inflate their metacomprehension, and (c) produce poorer mental visualizations of the system. We also tested the accuracy of performance standard (APS) hypothesis, which predicted that adding representational animation would enhance the accuracy of learners’ metacomprehension. Both the IU and the APS hypothesis predicted that directive animation would have no effect on metacognitive monitoring.

The results indicated that representational animation induced an illusion of understanding (i.e., decreased JOD and increased JOC) as predicted by the IU hypothesis; however, this illusion was induced only for low-proficiency learners (i.e., learners who had low overall learning scores). Representational animation did not have a statistically significant effect on JOC accuracy. Overall, the pattern of effect of representational animation and of directive animation on JOD and JOC indicated that adding animation to multimedia presentations affects metacognitive monitoring in complex ways that were not predicted by the IU or the APS hypothesis.

Learning Outcome

The negative effects of representational animation and the positive effects of directive animation on transfer learning found in the present study are consistent with prior studies (de Koning et al., 2009; Hegarty et al., 2003; Höfler & Leutner, 2007; Mayer et al., 2005; Schnotz & Lowe, 2005). The treatment conditions and the outcome measures used in this study were adapted from Experiment 2 of Mayer et al. In that study, the effect size of representational animation on transfer learning was Cohen’s $d = 1.06$. In the present study, the effect size was $d = 0.56$.

Illusion of Understanding

With respect to representational animation, the assertions of the IU hypothesis were partially supported. Representational animation induced an illusion of understanding (Betrancourt, 2005; Kühl et al., 2011; Lewalter, 2003; Schnotz & Rasch, 2005) but only among a subpopulation of participants with low overall learning scores. We refer to these learners as low-proficiency learners. Among low-proficiency learners, representational animation learners found the multimedia presentation easier and produced more optimistic metacomprehension. That is to say, low-proficiency representational animation learners experienced an illusion of understanding.

On the other hand, the overall pattern of effect of representational animation and of directive animation on JOD and JOC was more complex than predicted by the IU hypothesis. First, the effects of representational animation and of directive animation were moderated by learning proficiency. That is, there were attribute–treatment interactions (Cronbach & Snow, 1977) of $R \times L$ and of $D \times L$ on JOD and JOC. Second, the effects of directive animation were strikingly similar to the effects of representational animation. Indeed, the effects of directive animation were even more pronounced than those of representational animation. Directive animation induced two different forms of IU. Optimistic IU refers to lower JOD and higher JOC. Pessimistic IU refers to higher JOD and lower JOC. Directive animation induced optimistic IU in low-proficiency learners but pessimistic IU in high-proficiency learners. This bipolar effect of directive animation was consistent with expertise reversal effect (Plass, Kalyuga, & Leutner, 2010).

Accuracy of Performance Standard

With respect to representational animation, the results neither refuted nor supported the JOC accuracy assertion that representational animation enhances JOC accuracy. The difference in JOC accuracy between representational animation learners and no representational animation learners was not statistically significant. With respect to directive animation, the results contradicted the prediction of the APS hypothesis that directive animation would not affect JOC accuracy. Directive animation significantly degraded JOC accuracy.

Conclusion

We began this paper with the observation that representational animation has bipolar effects on learning. That is, representational animation can have a positive effect on learning under certain conditions but a negative effect under different conditions (Betrancourt, 2005; Höfler & Leutner, 2007; Tversky, & Morrison, 2002). Multimedia theorists offered the IU model as a metacognitive explanation for why representational animation has sometimes negatively affected learning (Betrancourt, 2005; Schnotz & Rasch, 2005). The results of this study showed that the IU model can explain not only the negative effects of representational animation but also the bipolar effects...
of animation in general. Learning is impeded when animation induces optimistic IU, and learning is enhanced when animation induces pessimistic IU.

Although the results showed that animation perturbs metacognitive monitoring (e.g., raises and lowers JOD and JOC), a causal explanation for this perturbation remains unresolved. Explanations that rely on the representational characteristics of animation (e.g., the IU and the APS hypotheses as defined in the introduction) do not explain the following: (a) Why were the effects of animation moderated by learning? (b) Why did direct animation affect JOD and JOC? (c) Why were the effects of representational and directive animations so similar to each other? However, given that the present study included one experiment using one set of content material (i.e., the workings of a flushing toilet tank), our findings should be replicated and extended to new materials.

References


The narration of the six segments of the multimedia presentation in the treatment was as follows:

1. A flushing toilet tank is made up of a number of parts. The tank and the lid store the water used to flush the toilet and house the other parts. The flow of water into the tank is controlled by these parts: the inlet pipe, the inlet valve, the inlet valve arm, the float arm, and the float. The rise and fall of the float pushes the inlet valve in and out of the inlet pipe. The flow of water out of the tank and into the toilet bowl is controlled by these parts: the handle, the connecting rods, the upper disc, the lower disc, the siphon bell, and the siphon pipe. The upper disc is free to move up and down on its own. The lower disc can be moved up and down with the handle.

2. Phase 1 - Starting the flush. When the handle is pressed down, the connecting rods are pulled up, causing the lower disk to rise and to push up the upper disk. As a result, the water in the siphon bell is forced over the siphon pipe into the toilet bowl.

3. Phase 2 - Continuing the flush. Once the handle is released, the two disks start to drop and separate from each other. As a result, the water flows through the holes in the lower disk, around the edges of the upper disk, over the siphon pipe, and into the toilet bowl. Note: the two disks separate, because the water that flows through the holes in the lower disk pushes up the upper disk.

4. Phase 3 - Starting the refill. As the water flows out of the tank, the water level drops. As the water level falls, the float drops toward the bottom, pulling out the inlet valve, and uncovering the hole in the inlet pipe. This allows the water to flow into the tank.

5. Phase 4 - Ending the flush. When the water flows out of the tank as well as into the tank, the water level continues to drop because the flow of water out of the tank is faster than into the tank. When the water level falls below the bottom of the siphon bell, air enters and breaks the siphon. This stops the flow of water into the toilet bowl.

6. Phase 5 - Ending the refill. When water flows into the tank but not out of the tank, the water level rises. As the water level rises, the float rises, pushing in the inlet valve, and closing the inlet hole. When the water level rises high enough, the flow of water into the tank is stopped. Now, the tank is ready for the next flush.

Received October 16, 2011
Revision received August 22, 2012
Accepted September 4, 2012