Picture this! Effects of photographs, diagrams, animations, and sketching on learning and beliefs about learning from a geoscience text

Jennifer Wiley

Department of Psychology, University of Illinois at Chicago, Chicago, Illinois

Correspondence
Jennifer Wiley, Department of Psychology, University of Illinois at Chicago, 1007 W. Harrison St. (MC 285), Chicago, IL 60607. Email: jwiley@uic.edu

Funding information
National Science Foundation, Grant/Award Number: DUE 1535299

Summary
Many studies have demonstrated that illustrating expository science texts with images that are interesting, but irrelevant for understanding the causal relations underlying scientific phenomena, can cause seduction effects, which can reduce understanding from text. The term “seduction effects” refers to the influence that images are thought to have on readers, seducing them away from deeply processing important information. The present study explores whether images relevant for instructional goals may also show some seduction effects. In this study, the presence of photographic images negatively impacted understanding compared with the presence of relevant animations or instructing students to sketch a drawing during reading. However, the results showed that both photographic images and relevant animations could lead to illusions of understanding, whereas sketching did not. The results suggest that even images that are relevant for instructional goals may sometimes result in seduction effects that deceive readers when judging their own understanding.

KEYWORDS
comprehension monitoring, expository text comprehension, metacomprehension, multimedia, spatial thinking

1 | INTRODUCTION

Open any science textbook from early grades through introductory college courses, and you are bound to see images accompanying the text. Images are added for a variety of reasons including to motivate readers, to spark reader interest, and to promote enjoyment. Images are also used to convey information visually that may be difficult to express in writing. This practice is consistent with theoretical assumptions and prior work showing that spatial thinking and visualizations can be supportive of better comprehension in science (Mayer, 1989; Sanchez & Wiley, 2010; Uttal & Cohen, 2012; Van Meter & Garner, 2005). Although these goals for incorporating images as part of expository science texts are intended to support better instructional outcomes, prior work in cognitive and educational psychology has demonstrated that some images may cause seduction effects that can undermine learning.

2 | EARLY RESEARCH ON SEDUCTION EFFECTS

The history of research on seduction effects can be traced to studies that first demonstrated the potential for adverse consequences in student learning when interesting facts and details were sprinkled into educational texts (Garner, Gillingham, & White, 1989; Hidi, Baird, & Hildyard, 1982). Based on their presumed effects on learners, these were referred to as “seductive details.” In 1997, Harp and Mayer extended the investigation of seductive details into a new direction by including “seductive images” along with interesting facts and details within an expository science text on lightning formation. These images were related to the topic in that they showed photographs of lightning strikes and lightning damage. Further, the photographs were interesting to readers. However, these images were irrelevant to understanding the causal process underlying lightning formation, which was the
instructional goal for this unit (Mayer, 1989). In contrast, the instructional goals were better captured by a set of visual explanatory summary images that highlighted important steps and relations in the causal process. The two types of “seductive” features (textual details and images) were presented together along with the explanatory images in the early work by Mayer and his colleagues (Harp & Mayer, 1997, 1998; Mayer, Heiser, & Lonn, 2001), and students in these conditions performed less well on tests of understanding following the unit than students who did not get these two types of seductive features. Consistent with the labels given to the details and images, the decrement in understanding was presumed to be due to these features seducing the reader away from deeply processing the relevant information.

Subsequent work has explored the effects of seductive images alone (Lenzner, Schnitz, & Müller, 2013; Sanchez & Wiley, 2006). The present study follows in this tradition of exploring the effects of images on understanding of scientific processes from reading an expository science text, independent of adding interesting details. In addition, although it is generally the case that only decorative images that are irrelevant for understanding causal relations have been considered as causing seduction effects, the present study pushes on this boundary and asks whether images relevant for understanding may also show seduction effects in some ways.

### 3 | EXPLANATIONS FOR SEDUCTION EFFECTS

Harp and Mayer (1998) delineated three classes of explanations for how these images and details may undermine comprehension: diversion, disruption, and distraction. These accounts were not intended to be mutually exclusive, yet some have received more or less support than the others when one considers the results of prior studies that have explored the effects of illustrations on learning from expository science texts that describe a process, system, or phenomenon such as lightning formation (Rey, 2012). The first possible explanation, diversion, suggests that the presence of an irrelevant image undermines learning by giving the reader a misconception about the true purpose of a passage. That is, the reader is misled by the additional information into misperceiving what is important in the text. This explanation may provide the best fit for poor learning that occurs when a text starts with an interesting-but-irrelevant initial image. A second potential mechanism is disruption, whereby the presence of additional information prevents the reader from building a coherent mental model from the text. This class of explanations may provide the best fit for irrelevant textual details paradigms where the interesting bits of information are sprinkled between important sentences within a passage.

The class of explanations that have the strongest support for explaining poor learning from text with seductive features are those related to the reader’s ability to pay attention to important information, to resist distraction, and to engage in the processing of essential material (Rey, 2014; Sanchez & Wiley, 2006). This class of explanations is represented by the distraction account as proposed by Harp and Mayer (1998). Several subsequent papers have described similar alternatives using different language (Park, Moreno, Seufert, & Brünken, 2011; Rey, 2014; Sanchez & Wiley, 2006), but ultimately, each variation within this class of explanations boils down to limited attentional mechanisms as being responsible for poor learning when interesting-but-irrelevant images are presented alongside causal, expository science texts. When the reader’s limited attention is seduced away from important causal information in a text, the reader is unable to engage in the processing of essential information, fails to construct a causal model from the text, and is unable to demonstrate understanding of the text on questions that ask about causal relations and processes. This line of reasoning is well supported by the results of the Sanchez and Wiley (2006) study that demonstrated that it is particularly readers who are low in working memory capacity who are harmed by interesting-but-irrelevant images.

In addition, recent work on comprehension monitoring suggests another potential explanation for how readers of expository science text may be affected by the presence of images. This alternative approach can be characterized as a deception account. Several studies have now demonstrated that the presence of images can impact judgments of knowing, beliefs about learning, perceptions of understanding, and readers’ ability to monitor their own learning (Cardwell, Lindsay, Förster, & Garry, 2017; Ikeda, Kitagami, Takahashi, Hattori, & Ito, 2013; Jaeger & Wiley, 2014; Lenzner et al., 2013; Serra & Dunlosky, 2010; Wiley, Sarmento, Griffin, & Hinze, 2017). If readers are deceived about their own level of understanding by the presence of images, then poor learning can result from readers spending too little time or too little effort in attempting to construct a mental model from a text. This recent work has suggested that the presence of images could prompt readers to experience illusions of understanding for any number of factors including relying on their interest in the image, how easy the image seems to be to understand, how easily information about the topic seems to come to mind, or due to heuristic beliefs about how the presence of images can aid their learning. This deception account prompts the question whether the presence of even relevant images (that depict information relevant to understanding a process, system, or phenomenon) may sometimes lead to seduction effects that undermine monitoring.

### 4 | OVERVIEW OF THE PRESENT STUDY

The present study employed three basic instructional conditions that have been used in prior experiments exploring seduction effects with images. Students were asked to learn about the causes of climate change and global warming from an illustrated expository text including a photographic image that did not contain information relevant for causal understanding of the phenomenon (photograph), an expository text with no visual adjunct (no-image), or an illustrated expository text including a static diagram containing information relevant for causal understanding the phenomenon (diagram). Both perceptions of understanding and actual understanding were assessed to allow for the computation of a measure of comprehension monitoring accuracy. Although seduction effects were expected for the photographic image, the main goal of this study was to test whether a relevant diagram might also be seen to have seductive effects that could...
undermine comprehension monitoring and lead to illusions of understanding. In addition, three other conditions were included in this study: a note-taking condition where students took notes from an expository text with no visual adjunct (note-taking), an animated diagram condition that included an animated version of the static diagram as part of the expository text (animation), and a sketching condition (sketching) where students were asked to sketch a drawing from an expository text with no visual adjunct. The motivation behind including these three other instructional conditions and detail on each is explicated in the next section, which reviews prior work offering suggestions as to how the presence of images might affect comprehension monitoring.

4.1 How the presence of images could affect comprehension monitoring

Several studies have now demonstrated that the mere presence of an image can impact readers’ perceptions of knowing, expectations of learning, or evaluations about their understanding of scientific processes (Cardwell et al., 2017; Ikeda et al., 2013; Lenzner et al., 2013; Serra & Dunlosky, 2010; Wiley, Sarmento et al., 2017). Recently, Cardwell et al. (2017) reported that simply presenting uninformative images (i.e., a picture of a rainbow) immediately before asking individuals to make evaluative ratings about the extent to which they understand a process (i.e., how or why rainbows form) led to higher judgments of perceived knowledge than when individuals made ratings that were not preceded by images. Other studies show that judgments of understanding may be impacted by the expectations that students hold about how images affect learning. In Serra and Dunlosky (2010), undergraduates were asked to rate how well they expected to understand a text on lightning even before reading. Readers were aware that the text would contain either photographs or explanatory diagrams, and this increased the magnitude of their judgments of understanding compared with a plain text (no-image) condition. The same differences in judgment magnitude due to image condition were also seen after reading. However, when actual understanding was examined on comprehension tests that followed the second judgment, test performance was only better for the explanatory diagram condition, but not the photograph condition, in comparison with the plain text (no-image) condition. Thus, understanding did not actually improve from the photographs, even though students expected it to. Because differences in ratings arose even before reading the text and were similar to differences that were seen after reading, these results suggest that students relied on a generalized multimedia heuristic rather than engaging in monitoring of their own comprehension in order to make their judgments. They made judgments consistent with their beliefs and expectations that any kind of image would help to improve their understanding. This led to a dissociation between their predictive judgments and actual learning.

In another recent paper, Lenzner et al. (2013) measured the effects of decorative and instructional pictures on middle school students’ perceptions of difficulty of learning from an expository science text about optics. The text was from the domain of physics and entitled “Light and Shadows.” The decorative pictures were aesthetically appealing photographs representing natural phenomena described in the text. The instructional pictures were diagrams illustrating important ideas and relationships among concepts described in the text. When the illustrated text contained decorative photographs, it led to reduced perceptions of difficulty compared with when the text was presented without decorative photographs. A similar pattern was seen for the instructional pictures. However, even though students perceived the text to be easier to understand with decorative photographs, it was only when instructional diagrams were present that understanding actually improved. Based on this prior work, one prediction for the present study is that one could expect individuals to give higher ratings of understanding in the presence of images even if they do not actually aid understanding. When some images do not actually improve understanding, then dissociations may be seen between perceptions and actual learning outcomes.

A paper by Ikeda et al. (2013) also reported dissociations between predicted and actual learning outcomes for different image types. The learning material for this study was an expository text on brain activity in patients with depression, which was either illustrated with photographic brain (functional Magnetic Resonance Image, fMRI) images or with bar graphs. Importantly, the bar graphs and brain images presented the same information about brain activity that was relevant for understanding the text. Ikeda et al. (2013) reported that students who had read the text illustrated with the photographic fMRI images thought they understood the text better than students who had read the text illustrated with bar graphs. Although judgments of understanding were higher for the fMRI image condition than the bar graph condition, learning outcomes did not actually differ across these conditions. Thus, the brain images were not actually more helpful than the bar graphs in terms of readers’ understanding of the text, and this led to a disconnect between the patterns for judgments and the patterns for actual learning.

Although the results of the Ikeda et al. (2013) study suggest that poor monitoring might be more likely when people are exposed to photographic images, Wiley, Sarmento et al. (2017) found that undergraduates seemed aware that photographic images would be unlikely to increase their understanding about biological processes. Instead, the images that were perceived to be most conducive to understanding were diagrams that contained information related to systems or processes. If readers have this kind of general expectation, then this suggests that individuals might actually give higher ratings of understanding when diagrams including relevant information are presented. The key determinant for monitoring accuracy then is whether any differences in predictive ratings of understanding due to the presence of different image types will be similarly reflected in measures of actual understanding. If measures of actual learning outcomes do not vary in the same manner as predictive ratings, then this would result in differences in monitoring accuracy across image conditions. If instructionally relevant images do lead to the highest judgments of understanding, this raises the possibility that illusions of understanding might be observed if actual understanding is not as strong as was predicted. This pattern of results would be consistent with the proposed deception account.

The main assumption behind the deception account is that readers may get an illusion of understanding as a result of exposure to instructional explanations and graphics, particularly when coupled with a failure to engage in more active processing of that information. There are
several lines of evidence from prior work that can be seen as consistent with this suggestion. For example, one line of prior research suggests that when students are presented with an explanation as part of a text, such as the inclusion of an analogical example to support understanding, this can actually undermine monitoring accuracy (Jaeger & Wiley, 2015; Wiley, Jaeger, Taylor, & Griffin, 2018). It is possible that some instructional supports may make the comprehension process feel too easy, which may pre-empt students from engaging in more active processing of the information. If so, then students may also adopt a passive processing mode when relevant diagrams provide them with a visual explanation for the process they are attempting to learn about (Jaeger & Wiley, 2015; Wiley et al., 2018). In fact, in Jaeger and Wiley (2014), although the presence of decorative images was seen to undermine comprehension monitoring (compared with a plain text, no-image condition), the presence of instructional diagrams did not improve comprehension monitoring (compared with a plain text, no-image condition) unless students were explicitly prompted to engage in explanation of the diagrams to support their understanding of the text.

### 4.2 The animation condition (animation)

Another area of prior research that suggests that instructionally relevant visualizations might have negative effects on monitoring is work on learning from animations. Because animations are generally more interesting and engaging for readers, and do some of the work of understanding causal relations for readers, it is possible that the presence of an animation may lead to inflated or inaccurate perceptions of understanding. Although the intent behind including dynamic features in visualizations is to support better student understanding, a number of researchers have suggested that animations can also have negative consequences for learning (Schnozt, Böckhler, & Grzondziel, 1999). Either because they attract attention or because of their apparent simplicity, animations could cause learners to fail to engage in deeper processing than static images (Bétrancourt, 2005; Lowe, 2003; Rieber, 1989; Schnozt & Rasch, 2005). Thus, when animated diagrams are presented with expository science text, this also could be predicted to have effects similar to the negative effects on monitoring accuracy that have been seen due to the inclusion of instructional analogies (Jaeger & Wiley, 2015; Wiley et al., 2018).

The results of studies using think-aloud protocols provide some support for this proposition. Both Lewalter (2003) and Kühl, Scheiter, Gerjets, and Gemballa (2011) tracked readers’ comments during reading of an illustrated expository science text. Lewalter (2003) observed that students who were exposed to animated diagrams made “positive monitoring” comments affirming their understanding while reading more often than students who read text illustrated with static diagrams. Yet, no differences were seen in actual learning outcomes. Kühl et al. (2011) found similar patterns. Students were more likely to make positive monitoring comments affirming their understanding from the text in the dynamic image condition than the static image condition, whereas no differences were seen in learning. However, although these findings suggest that students may experience illusions of understanding from dynamic images, the same patterns were not seen for perceptions of difficulty. When students were asked to judge the perceived difficulty of the learning conditions, no differences were seen between the two image conditions, although both were rated as less difficult than a plain text, no-image condition. Thus, although prior studies offer suggestions that seeing relevant images might undermine comprehension monitoring versus a plain text, no-image condition, because none of these prior studies have shown effects on monitoring accuracy, this motivates the present study.

Based on this prior work, a fourth condition was included in the present study in which readers were presented with animated versions of the carbon cycle diagrams. In this animation condition, the instructionally relevant diagram was animated with arrows showing the direction of carbon exchange.

### 4.3 The sketching condition (sketching)

In addition, because spatial thinking and visualization can be helpful for learning and developing an understanding of scientific phenomena, it was important to find a condition that might support spatial thinking without prompting readers into a passive, receptive rather than more active, generative mode of processing the text. Thus, in a fifth condition in the present study, rather than presenting readers with visualizations, they were prompted to draw sketches of systems or phenomena as they were reading the unillustrated, no-image text (Schleinschok, Eitel, & Scheiter, 2017). There is of course a general concern about drawing tasks that some students may feel ineffective at drawing or may lack drawing skills, which may undermine potential benefits of sketching (Jaeger, Velazquez, Dawdanow, & Shipley, 2018). On the other hand, a general advantage of drawing tasks is that they prompt generative, constructive processing and active integration of information by readers (Bobek & Tversky, 2016; Hall, Bailey, & Tillman, 1997; Scheiter, Schleinschok, & Ainsworth, 2017; Schwamborn, Mayer, Thillmann, Leopold, & Leutner, 2010; Van Meter & Garner, 2005). In contrast to drawing, presenting visualizations to readers may prompt information to be processed more passively or may lead to a stronger sense of fluency, ease of processing, accessibility of information in memory, reduced sense of difficulty, or similar subjective reactions. In contrast, learning activities that require generation have been shown to lead to more accurate comprehension monitoring (de Bruin, Thiede, Camp, & Redford, 2011; Griffin, Wiley, & Thiede, 2008, 2018; Jaeger & Wiley, 2014; Redford, Thiede, Wiley, & Griffin, 2011; Thiede, Anderson, & Therriault, 2003; Thiede, Dunlosky, Griffin, & Wiley, 2005; van Loon, de Bruin, van Gog, van Merriënboer, & Dunlosky, 2014).

### 4.4 The note-taking condition (note-taking)

Finally, as a control for the generative activity required by the sketching condition, in the present study, a sixth group of participants engaged in a note-taking condition. Students in this condition read the un-illustrated, no-image text and were prompted to take notes while reading. This was meant to promote generative, constructive processing as in the sketching condition but without the spatial thinking component.

These six conditions were used to test four hypotheses about whether the presence of relevant images may also be linked to seduction effects:
5.1 Participants

The sample was composed of 120 undergraduates (M = 21.13 [SD = 1.91] years of age, 77 females, 43 males) who participated in the study for course credit in psychology courses. Students were randomly assigned to the six conditions. Although the sample size is modest (most studies examining comprehension monitoring accuracy generally have 24–40 participants per cell), observed power was 0.94 for monitoring accuracy analyses. (Observed power was 0.45 for predictive judgment analyses and 0.98 for test score analyses.)

The conditions did not differ age, gender, scores on a college readiness test (ACT composite), or scores on the paper folding test (F ≤ 1). All persons gave informed consent prior to their inclusion in the study.

5.2 Materials and procedure

Participants read a textbook excerpt (~1,500 words) through a PowerPoint presentation about the Earth’s climate, global warming, and its relation to the carbon cycle and the greenhouse effect. The text was written at the seventh grade level, with a Flesch Reading Ease of 66.6, and was presented in six slides.

Students were randomly assigned to one of six conditions: illustrated text with a photograph depicting flooded city streets to emphasize that flooding is becoming more frequent with climate change (photograph), plain text (no-image), illustrated text with a static diagram of the carbon cycle (diagram), plain text with a prompt to take notes (note-taking), illustrated text with an animated version of the carbon cycle where arrows were animated to show the movement of carbon through the system (animation), or plain text with the prompt to “draw a representation of the earth’s climate system” (sketching). The photograph condition was intended to serve as an interesting-but-irrelevant-for-understanding image condition. For each image condition, the images were inset in the top right corner of Slides 4 and 5. Students read the PowerPoint slides at their own pace. Reading times were not collected.

After reading, students were asked to make a numerical judgment out of 5 in response to the prompt “How well do you understand the concepts in this text?” Then students were asked to predict their performance on an upcoming test of the material with 12 items in response to the prompt “How much do you understand about why global temperatures are changing? On a test of your understanding, how many items do think you will get correct out of 12?” For both numerical judgments, students were presented with a numerical scale (up to 5 or 12) with no other verbal anchors. They then wrote a short-answer response to the question “How and why recent patterns in global temperature are different from what has been observed in the past” and took a 12-item verification test on the material. The test questions represented inferences that could be made or not based on the text. Both measures were based on Jaeger and Wiley (2015).

After taking the verification test, students completed a demographic survey on which they rated their interest in the material (interestingness) and how much they felt that the material helped to support an understanding of “How and why recent patterns in global temperature are different from what has been observed in the past” (helpfulness) on a 10-point scale with the anchors “not very” to “very.” ACT composite scores were also collected. The ACT is a test of college readiness. Scores were collected to test for equivalence of conditions in general ability.

Finally, participants completed a computerized version of the paper folding test from the kit of factor-referenced cognitive tests (Ekstrom, French, Harman, & Dermen, 1976). In this test, the
participant must determine which one of five possible patterns of holes will result after a square piece of paper goes through a sequence of folds and then is punched. The test consists of 20 items, divided into two parts, presented one at a time. Participants had 3 min to complete each part. Participants’ scores were the total number of correct responses across both parts. This task was included because it is a standard aptitude measure often used in research on learning in STEM areas, which is important to equate across conditions (Höffler, 2010; Mayer & Sims, 1994; Sanchez & Wiley, 2010).

5.3 | Coding and computations of scores

The 12-item inference verification test was scored by giving 1 point for each correct answer. Total scores ranged from 3 to 12, with an average score of M = 8.59 (SD = 1.78). A converging measure of understanding was obtained by scoring responses to the short-answer test question (“How and why are recent patterns in global temperature different from what has been observed in the past?”) for the presence of five concepts following Jaeger and Wiley (2015). Two raters scored all short-answer responses with high interrater reliability (Krippendorf’s \( \alpha = 0.85, p < 0.05 \)). Total scores ranged from 0 to 5, with an average score of \( M = 1.89 (SD = 1.32) \). As in prior work (Griffin, Wiley, Britt, & Salas, 2012; Wiley, Hastings et al., 2017), scores on the inference verification test were significantly correlated with scores on the short-answer test (\( r = 0.40, p < 0.05 \)) demonstrating the convergent validity of two measures of understanding. To put measures on the same metric, short-answer test scores and inference test scores were converted to proportions. Given the correlations among the two test scores, a composite outcome measure was computed for use in the main analyses.

The two predictive judgments (judgments of understanding and test predictions) were also highly correlated (\( r = 0.76, p < 0.001 \)). Given the correlations among the two predictive judgments, to parallel the treatment of the test data, these were also converted to proportions, and a composite measure was computed for use in the main analyses. Although this seemed to be a straightforward way to handle the numerical judgment data in a manner that directly paralleled the computation of a composite outcome measure, it could be questioned whether it is valid to use the numerical rating of understanding out of 5 as a criterion against which to judge actual performance on the short-answer task. Although making a numerical prediction for performance on a 12-item test seems transparent for students, it is not clear that students understood what they were being asked to predict with the numerical judgment of understanding. To ensure that the results were not being driven by invalid judgments of understanding, a parallel set of analyses were conducted using just the 12-item test performance prediction. The alternative analyses showed the same results as those using the composite measure (which might be expected given the high correlation between the two sets of predictions). Due to the similarity in results, only the analyses using the parallel composite measures of predictions and outcomes are described in detail in the following sections.

Illusions of understanding were measured using confidence bias following earlier work (Glenberg, Wilkinson, & Epstein, 1982; Jaeger & Wiley, 2015). This was computed as the signed difference between the prediction composite and the outcome composite for each individual. Absolute error was also computed using the absolute difference between prediction composites and outcome composites for each individual. Because readers were generally overconfident in their predictions of understanding, these two measures were highly correlated (\( r = 0.91, p < 0.001 \)). However, because average overconfidence is technically a measure of directional bias within a sample and not degree of accuracy of individuals in the sample (see Griffin, Wiley, & Salas, 2013, for discussion), analyses using absolute error measures are used to show which conditions are associated with more accurate comprehension monitoring and less illusion of understanding.

6 | RESULTS

6.1 | Perceptions of understanding

Table 1 reports average proportionalized judgments of understanding and predictions for test performance in each condition. Composite prediction scores are shown in Figure 1. An analysis of variance with condition as a between-subjects factor and the composite prediction measure as the dependent variable showed no significant main effect of condition on perceptions of understanding, \( F(5, 114) = 1.30, p = 0.27 \).

6.2 | Test scores

Table 1 reports average proportionalized scores for both short-answer tests and inference verification tests for each condition. An analysis of variance with condition as a between-subjects factor and the composite outcome measure as the dependent variable demonstrated a significant main effect of condition on performance, \( F(5, 114) = 5.16, p < 0.001 \).

<table>
<thead>
<tr>
<th>Judge understanding</th>
<th>Predict performance</th>
<th>Short-answer test</th>
<th>Verification test</th>
<th>Confidence bias</th>
</tr>
</thead>
<tbody>
<tr>
<td>Photograph</td>
<td>0.74 (0.20)</td>
<td>0.67 (0.14)</td>
<td>0.23 (0.20)</td>
<td>0.63 (0.16)</td>
</tr>
<tr>
<td>No-image</td>
<td>0.75 (0.13)</td>
<td>0.63 (0.14)</td>
<td>0.31 (0.26)</td>
<td>0.68 (0.14)</td>
</tr>
<tr>
<td>Diagram</td>
<td>0.72 (0.19)</td>
<td>0.63 (0.15)</td>
<td>0.34 (0.21)</td>
<td>0.73 (0.14)</td>
</tr>
<tr>
<td>Note-taking</td>
<td>0.75 (0.16)</td>
<td>0.70 (0.18)</td>
<td>0.43 (0.24)</td>
<td>0.70 (0.14)</td>
</tr>
<tr>
<td>Animation</td>
<td>0.81 (0.16)</td>
<td>0.74 (0.16)</td>
<td>0.42 (0.29)</td>
<td>0.79 (0.12)</td>
</tr>
<tr>
<td>Sketching</td>
<td>0.70 (0.17)</td>
<td>0.63 (0.16)</td>
<td>0.55 (0.28)</td>
<td>0.78 (0.15)</td>
</tr>
</tbody>
</table>
A significant effect of condition on absolute error, with the pattern shown with the 95% confidence intervals, there was error differed from zero in all conditions except sketching. Consistent $p < 0.001$. In addition, the 95% confidence intervals show that absolute for average absolute error versus zero was significant, $p < 0.01$, partial $\eta^2 = 0.19$. As shown in Figure 1, the best understanding was demonstrated by readers in the sketching and animation conditions and the worst by readers in the photograph condition. Using Bonferroni corrections to adjust for multiple comparisons resulted in significant differences between these best and worst conditions. In addition, these adjusted comparisons also found that sketching led to better understanding than was seen in the no-image condition. No other differences were significant using adjusted comparisons.

### 6.3 Absolute error

Average absolute error in each condition is also shown in Figure 1. This measure represents the average of absolute differences between prediction composites and outcome composites. The one-sample $t$ test for average absolute error versus zero was significant, $t(119) = 11.33$, $p < 0.001$. In addition, the 95% confidence intervals show that absolute error differed from zero in all conditions except sketching. Consistent with the pattern shown with the 95% confidence intervals, there was a significant effect of condition on absolute error, $F(5, 114) = 4.67$, $p < 0.01$, partial $\eta^2 = 0.17$. As shown in Figure 1, the least absolute error was seen in the sketching condition. Using Bonferroni corrections to adjust for multiple comparisons resulted in significant differences between the sketching condition and three other conditions: photograph, no-image, and animation conditions. No other differences were significant using adjusted comparisons.

### 6.4 Confidence bias

Average confidence bias results were similar to those for absolute error. As shown in Table 1, readers across all conditions except sketching tended to be overconfident with respect to their judging and their understanding of the text. The one-sample $t$ test for average confidence bias versus zero was significant, $t(119) = 8.08$, $p < 0.001$.

There was a significant effect of condition on confidence bias, $F(5, 114) = 3.91$, $p < 0.01$, partial $\eta^2 = 0.15$. The most overconfidence was demonstrated by readers in the photograph condition and the least in the sketching condition. Using Bonferroni corrections to adjust for multiple comparisons resulted in significant differences between these best and worst conditions. In addition, adjusted comparisons also showed that sketching led to less overconfidence than was seen in the no-image condition. No other differences were significant using adjusted comparisons.

### 6.5 The role of interest and helpfulness

Figure 2 shows the average ratings of interest in the materials as well as the extent to which readers felt the materials were helpful for supporting their understanding. The effect of condition on ratings of interest did not reach significance, $F(5, 114) = 1.70$, $p = 0.14$. However, it is important to note that the photographic image was not perceived as being particularly interesting.

A significant effect of condition was found for ratings of helpfulness of the materials, $F(5, 114) = 2.63$, $p < 0.03$, partial $\eta^2 = 0.10$. As shown in Figure 2, the most helpfulness was perceived by the note-taking condition, whereas readers judged the photograph condition to be the least helpful. Using Bonferroni corrections to adjust for multiple comparisons resulted in significant differences between these best and worst conditions. No other differences were significant using adjusted comparisons.

### 6.6 what do readers use to predict their understanding?

Relations among the composite predictions, interest ratings, helpfulness ratings, and the composite from actual learning outcomes were examined. The correlations among these measures are presented in Table 2a. To understand which of these cues readers may be using as a basis for their judgments of understanding, three predictors (ratings of interest, ratings of helpfulness, and actual learning outcome composites) were entered into a simultaneous regression, $R^2 = 0.13$, $F(3, 116) = 5.86$, $p < 0.001$. As shown in Table 2b, ratings of interest were found to be the only significant predictor of the composite predictions. If one considers the outcome composite as indicative of actual understanding, then variance in the composite
predictions should be best explained by that measure. The fact that variance in the composite predictive judgments was best explained by interest ratings shows that the interest evoked by learning materials is contributing to readers’ evaluations of their own understanding.

6.7 | What predicts comprehension monitoring accuracy?

A second regression was used to provide a test of the subsidiary hypotheses that conditions that engaged readers in spatial thinking, involved generative learning activities, or evoked differences in readers’ interest would affect comprehension monitoring accuracy. To explore the spatial thinking hypothesis, a contrast compared the diagram, animation, and sketching conditions versus all others. To test the generative activity hypothesis, a contrast compared the sketching and note-taking conditions versus all others. The effect of interestingness was analyzed using the continuous ratings of interest. To understand which of these predictors had unique effects on monitoring accuracy, these three predictors were entered into a simultaneous regression, $R^2 = 0.18$, $F(3, 116) = 8.29$, $p < 0.001$. As shown in Table 3, all three factors were significant predictors of absolute error. Engaging in spatial thinking and engaging in generative learning activities each had a significant negative effect on the amount of error (i.e., they each reduced the amount of error in comprehension monitoring). On the other hand, higher interestingness ratings led to greater error in comprehension monitoring. Consistent with the earlier analyses, this regression approach suggests that the sketching condition showed the most accurate comprehension monitoring because readers engaged actively in relevant spatial thinking, whereas more error in comprehension monitoring was seen when readers engaged more passively with images and rated them as more interesting.

7 | DISCUSSION

In this study, actual learning outcomes were best in the conditions where readers either viewed instructionally relevant animations or generated sketches while learning, consistent with theoretical assumptions and prior work showing that spatial thinking and visualizations can be supportive of better comprehension in science (Mayer, 1989; Sanchez & Wiley, 2010; Uttal & Cohen, 2012; Van Meter & Garner, 2005). On the other hand, a contrast can be seen between these two instructional conditions in relation to their effects on monitoring accuracy. Readers who were exposed to the animated diagrams experienced illusions of understanding, whereas readers who were instructed to sketch a model of the earth’s climate system did not.

There was partial support for the image presence hypothesis. The fact that readers in the animated diagram condition had poor monitoring accuracy shows that in some cases relevant images can prompt illusions of comprehension. This result is consistent with prior suggestions that animations may seem particularly easy to process (Bétrancourt, 2005; Lowe, 2003; Rieber, 1989; Schnottz & Rasch, 2005). Further, this result is consistent with prior work suggesting that animations can undermine comprehension monitoring based on evidence from think-aloud comments (Kühl et al., 2011; Lewalter, 2003). The present study provides a novel demonstration in support of this suggestion by using a measure of monitoring accuracy computed from predictive judgments of understanding. However, it was not the case that the presence of any type of image was sufficient to undermine monitoring accuracy compared with the no-image conditions.

The sketching and note-taking conditions offer another important contrast, which provides support for the spatial thinking hypothesis. Although neither of these conditions presented images to readers, the sketching condition resulted in better learning and more conservative judgments, whereas the note-taking condition led to strong perceptions of helpfulness but did not result in improved understanding (relative to the no-image condition) to the same extent as the sketching condition. The benefits of instructing students to sketch a drawing while learning in science are consistent with prior work (Bobek & Tversky, 2016; Hall et al., 1997; Scheiter et al., 2017; Schwamborn et al., 2010; Van Meter & Garner, 2005). It interesting that a note-taking activity was not as effective, yet it seems that students may have the expectation that taking notes during learning will help their understanding. This novel finding that students may harbor expectations and beliefs about the effectiveness of note-taking activities will be important to explore in future work.

Some support was also found for the generative activity hypothesis. The benefits seen here resulting in lower absolute error for readers in the sketching condition are consistent with the results of Schleinschok et al. (2017) who found that students who engaged in a drawing activity were better able to monitor their understanding from a science text and made better restudy decisions. They are also consistent more broadly with other work that has found that

### TABLE 2A

Simple correlations among composite predictions, composite learning outcomes, and ratings of interest and helpfulness

<table>
<thead>
<tr>
<th>Predictions</th>
<th>Outcomes</th>
<th>Interest</th>
<th>Helpfulness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predictions</td>
<td>—</td>
<td>0.19</td>
<td>0.35*</td>
</tr>
<tr>
<td>Outcomes</td>
<td>—</td>
<td>0.15</td>
<td>0.14</td>
</tr>
<tr>
<td>Interest</td>
<td>—</td>
<td>0.58**</td>
<td></td>
</tr>
<tr>
<td>Helpfulness</td>
<td>—</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*p < 0.05. **p < 0.01.

### TABLE 2B

Regression analysis predicting composite predictions

<table>
<thead>
<tr>
<th></th>
<th>$\beta$</th>
<th>$t$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Composite learning outcomes</td>
<td>0.13</td>
<td>1.50</td>
<td>0.14</td>
</tr>
<tr>
<td>Ratings of helpfulness</td>
<td>−0.09</td>
<td>−0.82</td>
<td>0.41</td>
</tr>
<tr>
<td>Ratings of interest</td>
<td>0.36</td>
<td>3.47</td>
<td>0.001</td>
</tr>
</tbody>
</table>

### TABLE 3

Regression analysis predicting absolute error

<table>
<thead>
<tr>
<th></th>
<th>$\beta$</th>
<th>$t$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spatial thinking</td>
<td>−0.27</td>
<td>−3.14</td>
<td>0.002</td>
</tr>
<tr>
<td>Generative activity</td>
<td>−0.27</td>
<td>−3.23</td>
<td>0.002</td>
</tr>
<tr>
<td>Interestingness ratings</td>
<td>0.19</td>
<td>2.20</td>
<td>0.03</td>
</tr>
</tbody>
</table>

Note. Positive predictors mean more error and less monitoring accuracy, and negative predictors mean less error and better monitoring accuracy.
generative learning activities can help to improve comprehension monitoring (de Bruin et al., 2011; Griffin et al., 2008, 2018; Jaeger & Wiley, 2014; Redford et al., 2011; Thiede et al., 2003, 2005; van Loon et al., 2014). Further, these results suggest that by combining spatial thinking with a generative activity, the sketching condition was able to support the best monitoring accuracy from this expository science text. Although another recent study by Jaeger et al. (2018) found that a sketching activity was unable to reduce the effect of seductive textual details on comprehension, a next step for the present line of research would be to test if a sketching activity might instead help students to avoid the seductive effects of images (including photographic images and animations) on monitoring.

Finally, several results were relevant for the interest hypothesis in terms of the relations that were seen between interest and perceptions of understanding. On one hand, the results suggest that it was not heightened interest in the photographic image that led to the detriments in monitoring accuracy in that condition. The presence of photographs did not seem to alter predictive judgments in this study, as they did not differ from predictions in the no-image condition. The photographic image condition also did not appear to be more interesting than any other condition, and yet, it still harmed understanding. Similar to the results of Wiley, Sarmento et al. (2017), the college students in this study also did not seem to hold expectations that the decorative photograph would improve their understanding, as this image was rated lowest in helpfulness for understanding. Thus, neither a strong feeling of interest nor a perception of helpfulness appears to be related to the poor monitoring that was seen in the uninformative photographic image condition.

On the other hand, the first regression analysis showed that interest ratings were the only significant predictor of readers’ perceptions of understanding. This is consistent with the hypothesis that the interestingness of materials can serve a heuristic cue that readers may rely on when attempting to monitor their comprehension of a text (Griffin, Jee, & Wiley, 2009; Jaeger & Wiley, 2014; Lin & Zabrucky, 1998). In addition, the second regression analysis was able to show that as interest in the materials increased, this led to more error in monitoring. When readers are using subjective feelings such as interest, fluency, perceived ease, or expectations of helpfulness as the basis for their judgments of understanding, then this can explain poor calibration between predictive judgments and actual outcomes.

One limitation of the present study was that the sample sizes were relatively small in each condition. As has been the case in previous research on subjective perceptions of understanding, often no significant differences are seen in the magnitudes of judgments across conditions. In the present case, the relatively small sample resulted in low power for this analysis in particular. In the future, studies with larger samples will be needed to explore the possible effects of different instructional conditions on judgments of understanding more robustly.

In summary, this study provides several results that are consistent with a deception account. The deception account pushes on the boundary of what has traditionally been considered as “seduction” effect when learning from expository science texts. The present results demonstrate that there may be negative consequences even when a text is illustrated with images that are relevant for the instructional goal. The essence of the deception account is that the presence of even instructionally relevant images can impact readers’ ability to monitor their understanding. If readers are deceived about their own level of understanding and are unable to accurately monitor their comprehension in the presence of images, then poor learning could result from readers investing too little effort toward constructing a mental model from a text. The results of the present study show that lessons containing animations may be most at risk for prompting readers to have illusions of comprehension and can lead to seduction effects that undermine monitoring. In contrast, readers who were instructed to sketch a drawing during reading did not experience the same deception. The results suggest that the sketching condition showed the most accurate comprehension monitoring because readers were prompted to engage in a generative, spatial thinking activity, whereas animations led to more error in comprehension monitoring, and more deception as readers engaged more passively and rated the images as more interesting.

HUMAN STUDIES AND SUBJECTS

This research was performed under a protocol approved by the Institutional Review Board of the UIC Office for the Protection of Research Subjects.

ACKNOWLEDGMENTS

This work was supported by the National Science Foundation under DUE Grant 1535299. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation. The author thanks Thomas D. Griffin, Keith Thiede, Allison Jaeger, Tara Brown, Jennifer Chun, Tim George, Tricia Guerrero, Morgan Hager, Marta Mielicki, Mayte Noreiga, Rachel Parker, Chelsea Perschon, David Sarmento, and Uliana Solovieva for discussion and assistance on this project.

CONFLICT OF INTEREST

There are no conflicts of interest to declare.

ORCID

Jennifer Wiley © https://orcid.org/0000-0002-2590-7392

REFERENCES


**How to cite this article:** Wiley J. Picture this! Effects of photographs, diagrams, animations, and sketching on learning and beliefs about learning from a geoscience text. *Appl Cognit Psychol. 2018:1–11*. https://doi.org/10.1002/acp.3495